Jamie Berryhill
Kévin Kok Heang
Rob Clogher
Keegan McBride

HOLA, MUNDO: LA INTELIGENCIA ARTIFICIAL Y SU USO EN EL SECTOR PÚBLICO

DOCUMENTOS DE TRABAJO DE LA OCDE SOBRE GOBERNANZA PÚBLICA NÚM. 36

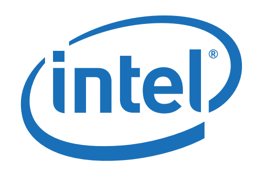
https://dx.doi.org/10.1787/726fd39d-en

Traducido por





Traducción patrocinada por



Publicado originalmente por la OCDE en inglés con el título: "Hello, World: Artificial intelligence and its use in the public sector", OECD Working Papers on Public Governance, No. 36 © OECD 2019, https://doi.org/10.1787/726fd39d-en

Imagen de portada: Freepik.com (vector creado por iuriimotov).

Presentación a la edición en español

Ante el avance acelerado de la tecnología, el cambio climático, la globalización y, más recientemente, la pandemia de COVID-19, la comunidad internacional se enfrenta a nuevos paradigmas que requieren de un cambio sistémico que apunte a la resiliencia global y a la construcción de mecanismos de gobernanza y la búsqueda de consenso para afrontarlos.

Actualmente nos encontramos en un punto de convergencia de varias tecnologías: la inteligencia artificial (IA), el Internet de las cosas (IoT) y la transformación de las redes 5G, que, en conjunto, definirán el futuro de la industria de la información y las telecomunicaciones donde los datos emergen como una fuerza de transformación en una época en la que el auge de los dispositivos define en gran medida nuestras interacciones sociales, comerciales, financieras, laborales, educativas y con el gobierno.

Por este motivo, es un honor para Intel poder contribuir al trabajo de la OCDE y poner al alcance del mundo de habla hispana la traducción de este importante documento de trabajo intitulado *Hola, mundo: La inteligencia artificial y su uso en el sector público*, dirigido a funcionarios de gobierno con el afán de resaltar la importancia de la IA y de sus aplicaciones prácticas en el ámbito gubernamental.

En este manual, además de compartir enfoques desarrollados por diversos gobiernos, también aborda cómo los responsables de política pública y los funcionarios gubernamentales pueden definir una estrategia para maximizar los beneficios de la IA al explorar en la ciencia de datos como una herramienta para incrementar la productividad del servicio público, servir mejor a sus ciudadanos y potenciar la innovación en sus empresas.

En este contexto, el papel del Estado es determinante para transitar a una agenda nacional que se incorpore a los planes, programas y presupuestos del Gobierno Federal en donde se fomente también el desarrollo regional, la formación de capital humano, el emprendimiento y esquemas de inversión pública y privada bajo un enfoque multisectorial.

En 2020, Intel, tanto de manera independiente como en colaboración con la Asociación de Internet, ha contribuido a elevar el nivel de la conversación sobre la IA entre los principales actores que forman parte del ecosistema sobre la importancia estratégica del impulso y desarrollo de la IA.

Cómo se beneficien México y otros países de la revolución tecnológica dependerá en gran medida de la existencia de un plan de acción que sea transversal, visionario, inclusivo y pragmático; uno que maximice los beneficios de las nuevas tecnologías, minimice los riesgos y priorice que la innovación es una herramienta medular para garantizar el bienestar social. Un plan incluyente que pueda ser implementado por los distintos órdenes de gobierno y por los actores relevantes de cada sector. Asimismo, alinear la narrativa con

acciones puntuales y apostar al poder de las colaboraciones, bajo un enfoque de múltiples actores tanto en el país como en el ámbito regional e internacional.

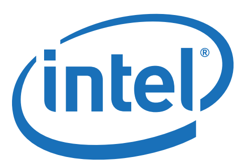
Finalmente, en Intel no solo creamos tecnología que cambia el mundo, habilita el progreso y mejora la vida de las personas, también aspiramos a crear, innovar y poner a prueba los límites de la tecnología en un marco ético. Deseamos influenciar a la industria en materia de responsabilidad social, sostenibilidad, diversidad e inclusión y trabajar de manera conjunta con los distintos actores de gobierno, la academia y la sociedad en general para hacer frente a los problemas que solo juntos podremos solucionar.

Ciudad de México, diciembre de 2020

Traducido por



Traducción patrocinada por



Prólogo

La Inteligencia Artificial (IA) es un área de investigación y aplicación tecnológica que puede tener un impacto significativo de muchas maneras sobre las políticas y servicios públicos. Se espera que en pocos años exista la posibilidad de liberar casi un tercio del tiempo de los funcionarios públicos, permitiéndoles pasar de actividades mundanas a trabajo de alto valor. Los gobiernos también pueden utilizar la IA para diseñar mejores políticas y tomar mejores decisiones, mejorar la comunicación y el compromiso con los ciudadanos y residentes, así como mejorar la velocidad y la calidad de los servicios públicos. Si bien los posibles beneficios de la IA son importantes, alcanzarlos no es una tarea fácil. El uso de la IA por parte del gobierno sigue el mismo camino que el del sector privado; el área es compleja y tiene una curva de aprendizaje pronunciada; y el propósito y contexto gubernamentales son únicos y plantean una serie de desafíos.

El Observatorio de Innovación en el Sector Público (*Observatory of Public Sector Innovation*, OPSI, por sus siglas en inglés) (https://oecd-opsi.org) de la OCDE elaboró este documento de trabajo, *Hola, mundo: la Inteligencia Artificial y su uso en el sector público,* para ayudar a los funcionarios del gobierno a que comprendan la IA y exploren cuestiones específicas del sector público. "¡Hola, mundo!" es tradicionalmente el primer programa informático creado por alguien que está aprendiendo a codificar y este manual básico está dirigido a los funcionarios públicos para que puedan dar sus primeros pasos hacia la exploración de la IA. Se basa en la labor del proyecto Transición Digital (*Going Digital*) de la OCDE, un futuro Observatorio de Políticas en materia de IA (*AI Policy Observatory*) de la OCDE (http://oecd.ai) y los Líderes de Gobierno Electrónico (*E-Leaders*). Es el segundo de una serie de resúmenes sobre temas de interés para la comunidad de innovación en el sector público, después de *Blockchains Unchained*, publicado en junio de 2018.

En un momento de creciente complejidad, incertidumbre y demandas cambiantes, los gobiernos y los funcionarios públicos necesitan comprender, evaluar e incorporar nuevas formas de hacer las cosas. El OPSI les ayuda arrojando luz sobre los esfuerzos de los gobiernos por crear políticas y servicios más eficientes y adecuados y acompañándolos al momento de explorar e implementar enfoques innovadores. El gobierno puede utilizar la IA para innovar, de hecho, en muchos casos ya lo está haciendo. Por ejemplo, varios líderes a nivel mundial ya cuentan con estrategias para desarrollar la capacidad de la IA como prioridad nacional. La IA se puede utilizar para que los procesos actuales sean más eficientes y precisos. Se puede utilizar para consumir y analizar información no estructurada, como los tweets, para ayudar a los gobiernos a conocer las opiniones de los ciudadanos. Por último, al mirar hacia el futuro, será importante considerar y prepararse para las implicaciones de la IA en la sociedad, el trabajo y el propósito humano.

Agradecimientos

Hola, mundo: la Inteligencia Artificial y su uso en el sector público fue elaborado por la Dirección de Gobernanza Pública (GOV), bajo la dirección de Marcos Bonturi.

El manual fue elaborado por el Observatorio de Innovación en el Sector Público (OPSI), en colaboración con los Líderes de Gobierno Electrónico (*E-Leaders*) y la Red de puntos de contacto nacionales del observatorio (*Network of the Observatory National Contact Points*).

El manual fue elaborado por Kévin Kok Heang y Jamie Berryhill del OPSI; Rob Clogher, candidato a la Maestría Ejecutiva en Administración Pública en la Universidad de Nueva York y el Colegio universitario de Londres; y Keegan McBride, Gerente de GovAiLab en la Escuela de Tecnologías de la Información de la Universidad Tecnológica de Tallin. Piret Tõnurist y Alex Roberts del OPSI hicieron importantes contribuciones a la sección del Capítulo 4 referente al enfoque en el futuro y la innovación anticipada. El trabajo se llevó a cabo bajo la coordinación de Marco Daglio (Jefe del OPSI y Jefe Interino de División, RPS). Los colegas de la OCDE, entre ellos Luis Aranda, Barbara Ubaldi, Natalia Nolan Flecha, Delphine Moretti, Alistair Nolan y Karine Perset, lo revisaron e hicieron observaciones. Liv Gaunt y David McDonald brindaron asistencia editorial.

Asimismo, el equipo de la OPSI desea agradecer las contribuciones de varias partes interesadas que compartieron sus conocimientos mediante entrevistas, debates y correspondencia. En particular, el equipo agradece a Coline Cuau y Wietse Van Ransbeeck de CitizenLab; Dietmar Gattwinkel, Georges Lobo y Fidel Santiago de la Comisión Europea; Gregg Blakely, Ashley Casovan, Noel Corriveau, Benoit Deshaies, Chelsea Escott, Russel Gauthier, Cezary Gesikowski, Hubert Laferriere, Michael Karlin, Laura MacDonald, Stan Martens, Patrick McEvenue, Amanda McPherson, Mark Robbins y Jeremiah Stanghini del Gobierno de Canadá; Aleksi Kopponen y Niko Ruostetsaari, del Gobierno de Finlandia; Enzo Maria Le Fevre, del Gobierno de Italia; Kenji Hiramoto, Ken Tamaru y Hiroki Yoshida, del Gobierno de Japón; Farah Hussain y Sebastien Krier, del Gobierno de Reino Unido; Dan Chenok, de IBM Center for The Business of Government; Olivia Elson, de Results for Development; y Cosmina Dorobantu, Pauline Kinniburgh y Florian Ostmann, del Instituto Alan Turing.

Por último, el equipo desea agradecer las contribuciones de los particulares durante la fase de consulta pública de este manual (del 1 de agosto al 15 de septiembre de 2019). El equipo del OPSI ha realizado importantes revisiones y mejoras al informe con base en la retroalimentación recibida y los autores agradecen sinceramente a todos los que participaron. Aunque el OPSI no puede enumerar a todos los que participaron, cabe destacar

la retroalimentación de John Atkinson, Javier Barreiro, Sonia Castro, Lequanne Collins-Bacchus, Stefan Bergheim, Martine Delannoy, Shachee Doshi, Michael Greenwood, Alex Goncharov, Raed Mansour, Mind Senses Global, Maria Marques, Roland Pihlakas, Karmen Kern Pipan, la Asociación Portuguesa de Psicólogos, Olivia Shen, Craig Thomler, Stefan Torges, Colin van Noordt, Samuel Witherspoon y Nathan Young.

Índice

Resúmen ejecutivo

La Inteligencia Artificial (IA) es muy prometedora para el sector público y los gobiernos se encuentran en una posición única en relación con la IA. Son capaces de establecer prioridades, inversiones y reglamentos nacionales para la IA y también pueden utilizarla para redefinir las formas en que el sector público crea políticas y servicios. El despliegue publicitario en torno a las tecnologías emergentes con frecuencia exagera u oscurece las aplicaciones prácticas. Por lo tanto, comprender la IA es fundamental para ayudar a los encargados de formular políticas y a los funcionarios públicos a determinar si les puede ayudar a cumplir sus misiones.

Los particulares y las empresas interactúan cada vez más con la IA. Aunque este ámbito se ha investigado y debatido durante más de 70 años, aún no existe una definición aceptada a nivel general; la IA significa cosas diferentes para diferentes personas. De acuerdo con la Recomendación de la OCDE de 2019 sobre la inteligencia artificial, hoy en día la IA se refiere a los sistemas basados en máquinas que pueden, para un determinado conjunto de objetivos definidos por el ser humano, hacer predicciones, recomendaciones o decisiones que influyen en entornos reales o virtuales. El Observatorio de Innovación en el Sector Público (OPSI) de la OCDE (https://oecd-opsi.org) elaboró este manual básico para ayudar a determinar lo que significa la IA en términos de innovación en el sector público y para ayudar a los funcionarios públicos a comprender la IA y a explorar sus repercusiones en las políticas y los servicios.

A nivel técnico, si bien la IA adquiere una variedad de formas, toda la IA se puede clasificar actualmente como “IA estrecha”. En otras palabras, se puede utilizar para tareas específicas, por ejemplo, para manipular e interpretar textos mediante el procesamiento natural del lenguaje, detectar y clasificar objetos mediante la visión artificial y reconocer e interpretar lenguaje hablado y traducirlo a texto mediante el reconocimiento de sonidos vocales. Enfoques como el “aprendizaje no supervisado”, “aprendizaje supervisado”, “aprendizaje por refuerzo” y “aprendizaje profundo”, pertenecientes al campo del “aprendizaje automático” (*machine learning*), tienen un potencial significativo para diversas tareas, si bien cada uno tiene sus propios puntos fuertes y limitaciones. Sin embargo, es importante señalar que todo proyecto de IA parte de lo mismo: los datos. Los gobiernos deben asegurarse de tener acceso a datos suficientes, de calidad e imparciales antes de que puedan aprovechar estas técnicas de forma plena y ética.

En cuestión de la adopción de la IA, el sector público se ha quedado a la zaga del sector privado. Sin embargo, los gobiernos están tratando de darles alcance rápidamente. Para catalizar la innovación impulsada por la IA, a través de un mapeo inicial que realizó la OCDE

se identificó a 50 países (incluyendo la Unión Europea) que han lanzado, o tienen planes de lanzar, estrategias nacionales de IA. Si bien se encuentran en diferentes etapas de desarrollo, dichas estrategias incluyen algunos temas comunes: el desarrollo económico, la confianza y la ética, la seguridad y la mejora del procesamiento de talentos. De estos 50 países, 36 han desarrollado (o planean desarrollar) estrategias individuales para la IA del sector público o un enfoque específico integrado en una estrategia más amplia. Esto es fundamental, ya que permite que se integre a la IA en todo el proceso de formulación de políticas y diseño de servicios. Con frecuencia, estos componentes del sector público promueven una serie de temas comunes, por ejemplo:

- la experimentación y a algunas veces la financiación de la IA del gobierno para automatizar los procesos, guiar la toma de decisiones y desarrollar servicios anticipatorios para los ciudadanos.
- la colaboración intergubernamental, intersectorial e internacional a través de consejos, redes, comunidades y asociaciones.
- la gestión estratégica y el uso de los datos gubernamentales, incluyendo los datos abiertos, para alimentar a la IA en todos los sectores.
- el establecimiento de condiciones y lineamientos para el uso transparente, ético y confiable de la IA en el gobierno.
- el mejoramiento de la capacidad de la administración pública mediante la capacitación, las herramientas y la contratación.

Muchos gobiernos también han lanzado proyectos que utilizan la IA para mejorar la eficiencia y la toma de decisiones, fomentar relaciones positivas con los ciudadanos y las empresas, ayudar a alcanzar los Objetivos de Desarrollo Sostenible y resolver problemas en áreas críticas como la salud, el transporte y la seguridad. Por ejemplo, la iniciativa "bomba en una caja" (*bomb-in-a-box*) de Canadá utiliza la IA para ayudar a identificar la carga aérea de alto riesgo, y el asistente virtual 24/7 de Letonia denominado UNA, atiende preguntas de los clientes. Estos y otros proyectos e iniciativas estratégicas se presentan a manera de ejemplos y casos de estudio en este manual.

Si bien la IA *puede* ayudar a promover la innovación en el gobierno, no es la solución para todos los problemas. En general, los gobiernos deberían determinar si la IA es la mejor solución para un problema determinado mediante un análisis de las alternativas y los sacrificios, todo ello con base en una sólida comprensión de las necesidades de sus usuarios. Los gobiernos también deben considerar muchos aspectos al tratar de seguir explorando la IA. Por ejemplo, deberían:

- Brindar apoyo y una dirección clara, así como crear un espacio para la flexibilidad y la experimentación, por ejemplo, estableciendo estrategias y principios generales, expresando el apoyo del personal senior en torno a la experimentación de la IA y desarrollar estructuras dentro del gobierno para considerar nuevos enfoques y llevar los éxitos a mayor escala.
- Proporcionar perspectivas multidisciplinarias, diversas e inclusivas, por ejemplo, mediante la formación de equipos con diferentes habilidades, antecedentes, razas y géneros.

- Desarrollar un enfoque confiable, justo y responsable para el uso de la IA, por ejemplo, mediante el establecimiento de marcos jurídicos y éticos, prestando atención a las personas que pudieran verse afectadas, definiendo la función de los seres humanos en los procesos impulsados por la IA, buscando explicaciones de los resultados de la IA y desarrollando estructuras abiertas de rendición de cuentas.
- Garantizar la recopilación, el acceso y uso éticos de los datos de calidad, por ejemplo, utilizando estrategias para administrar los datos que permitan la legibilidad de estos en la máquina, promuevan su confidencialidad y seguridad y mitiguen los sesgos a lo largo de su ciclo de vida.
- Garantizar que las organizaciones gubernamentales tengan acceso a financiamiento, capacidad interna y externa, e infraestructura para utilizar la IA a través de la capacitación y reclutamiento, colaborando y asociándose a nivel externo, diseñando mecanismos de adquisición que funcionen para la IA y considerando las necesidades de infraestructura.
- Reconocer los posibles cambios importantes que la IA podría traer en el futuro y utilizar un enfoque previsor de innovación para explorar y dar forma de manera sistemática y dinámica al potencial de la IA en el sector público mientras aún sea posible.

Puede parecer abrumador la cantidad de consideraciones que los funcionarios públicos deben tener en cuenta. Sin embargo, los gobiernos de todo el mundo han ideado enfoques para abordarlas en su propio contexto. Este manual analiza muchos de estos enfoques, los cuales podrían adaptarse para que se utilicen en otros países y contextos.

Capítulo 1

Inteligencia Artificial: Definiciones y contexto

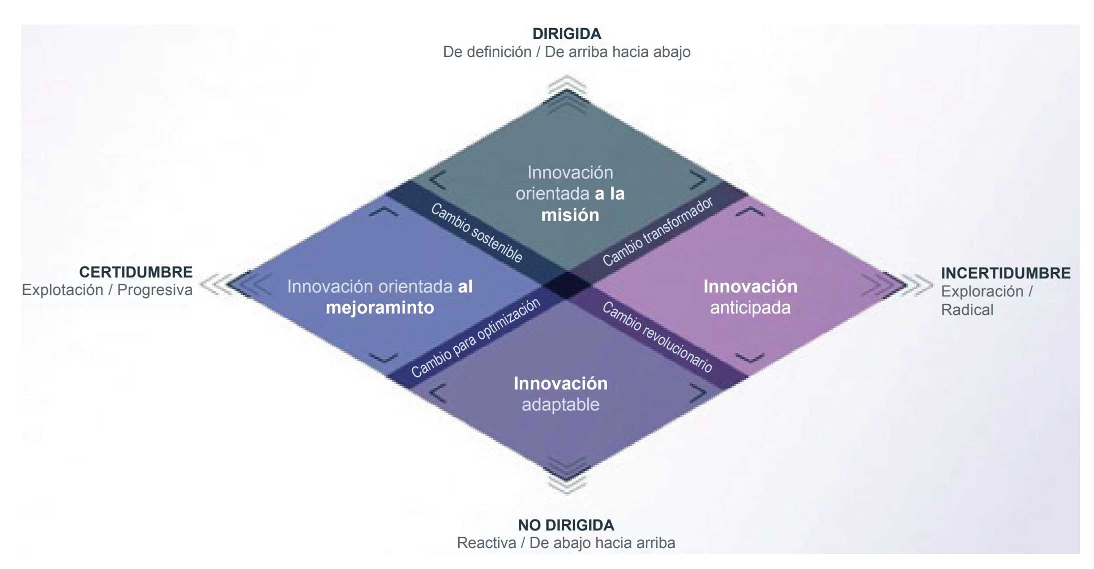
El Observatorio de Innovación en el Sector Público (*Observatory of Public Sector Innovation*, OPSI)[1] de la OCDE colabora con los gobiernos y los funcionarios públicos para:

- **Descubrir prácticas emergentes y determinar cuál es el siguiente paso**, identificando nuevas prácticas en la vanguardia del gobierno, conectando a aquellos que se involucran en nuevas formas de pensar y tomando en cuenta lo que significan los nuevos enfoques para el gobierno.
- **Explorar cómo convertir lo nuevo en normal**, estudiando la innovación en diferentes contextos públicos e investigando métodos para liberar la creatividad e incorporarla al trabajo de los funcionarios públicos.
- **Brindar asesoramiento confiable sobre cómo fomentar la innovación**, compartiendo orientación y recursos sobre las formas en que los gobiernos pueden apoyar la innovación para obtener mejores resultados.

Gracias a su labor con países de todo el mundo, el OPSI ha aprendido que la innovación no es una sola cosa, sino que adopta diferentes formas, todas las cuales deberían considerarse en el sector público. El OPSI ha identificado cuatro facetas de la innovación en el sector público (Figura 1.1)[2]

[1] https://oecd-opsi.org.
[2] Véase https://oecd-opsi.org/projects/innovation-facets para obtener más información sobre las facetas.

Figura 1.1: Cuatro facetas de la innovación en el sector público



Fuente: OPSI.

- La **innovación orientada a la misión** establece un resultado claro y un objetivo general para lograr una misión específica.
- La **innovación orientada al mejoramiento** actualiza las prácticas, logra eficiencias y mejores resultados y se basa en las estructuras existentes.
- La **innovación adaptable** analiza y pone a prueba nuevos enfoques a fin de responder a un entorno operativo cambiante.
- La **innovación anticipada** explora y se involucra en problemas emergentes que podrían determinar las prioridades y compromisos futuros.

A través de este documento, el OPSI ha descubierto que el enfoque óptimo es el enfoque de cartera para la innovación, el cual toma en cuenta una combinación de facetas de la innovación. En un mundo complejo, confiar en un solo enfoque es muy arriesgado. Por lo tanto, se debe disponer de varias opciones para compensar el riesgo y garantizar alternativas viables. El gobierno puede utilizar la IA, y en muchos casos ya lo está haciendo, para innovar de manera que se atraviesen por la ruta más corta todas las facetas de la innovación en el sector público. El OPSI ha elaborado este manual para ayudar a que los gobiernos logren un enfoque de cartera para la innovación, aprendiendo de las prácticas emergentes en el área, minimizando a la vez las consecuencias negativas e indeseables. Para ayudar aún más a los líderes e innovadores del sector público a hacerse camino a través del panorama de la IA y a conocer y, según corresponda, adoptar la IA, el OPSI ha desarrollado dos complementos digitales para el manual:

- **Estrategias de IA y componentes del sector público** (https://oe.cd/aistrategies): En él se documenta cómo los países están desarrollando estrategias de IA y la medida en que prevén específicamente la transformación del sector público. El OPSI planea integrar en el futuro este recurso con el[3] próximo depósito de conocimientos del Observatorio de

[3] http://oecd.ai.

Políticas en materia de IA (*AI Policy Observatory*) de la OCDE sobre estrategias nacionales de IA para que los usuarios puedan obtener periódicamente información completa y actualizada sobre éstas.

- **Herramientas y recursos de la IA** (https://oe.cd/airesources)**:** Proporciona un depósito categorizado de herramientas y recursos prácticos que pueden ayudar a los funcionarios públicos a aprender más sobre la IA y sus posibles funciones en el sector público.

Tal y como se describe en este manual, es evidente que la IA tiene un enorme potencial para la innovación en todos los sectores e industrias, tanto hoy en día como en el futuro. Debido a la creciente atención que se ha prestado a la IA en los últimos años, en ocasiones puede parecer que ésta irrumpió en escena apenas hace poco tiempo. Sin embargo, este tema se ha debatido e investigado durante más de 70 años. Hoy en día, la IA se puede encontrar en innumerables tecnologías y servicios: los algoritmos que las aplicaciones de navegación utilizan para evitar el tráfico, que Netflix y Spotify utilizan para recomendar películas y canciones, y que los proveedores de correo electrónico utilizan para filtrar automáticamente el spam, todos se basan en la IA. La IA ha sido objeto de debate a nivel general y se ha convertido en un activo fundamental en todos los sectores, incluyendo el gobierno. Docenas de países han desarrollado estrategias nacionales para la IA, y muchos se han comprometido a donar millones de euros (o su equivalente) para financiar la investigación y desarrollo, incluyendo el uso de la IA para que las operaciones gubernamentales sean más eficientes y respondan mejor a las necesidades de los ciudadanos y las empresas. En la actualidad, los gobiernos y sus socios en la industria, así como la sociedad civil están utilizando la IA para impulsar la innovación en áreas del sector público como el sector de la salud, el transporte y los Objetivos de Desarrollo Sostenible (ODS), tal como se describe en el Capítulo 3.

Pero, ¿qué *es* la IA?

Las representaciones de la IA en el cine o la televisión se prestan a transmitir visiones de supermáquinas poderosas, similares a los humanos, como el cíborg de *The Terminator* o sistemas informáticos conscientes con programación de varios niveles de calidad, como HAL 9000 de *2001: Una odisea del espacio*. En el mundo real, las expectativas de la gente sobre la IA van desde el entusiasmo a la ambivalencia y desde el optimismo al miedo. Aunque las inexactitudes científicas o la hipérbole en la ficción pueden jugar un papel, este amplio abanico de opiniones se debe también al hecho de que la IA significa cosas diferentes para diferentes personas. Hoy en día, aún no existe ninguna definición universalmente aceptada para la IA y es probable que no la haya pronto. Este manual no pretende resolver este problema, sino brindar algunos fundamentos sobre la naturaleza de la IA para ayudar a los funcionarios públicos a hacerse camino a través de este complejo terreno, distinguir entre la exageración y la realidad y estar mejor informados sobre lo que la IA puede significar en su propio contexto.

Definición de la Inteligencia Artificial

Cuando se habla de la "Inteligencia Artificial", puede ser difícil formular una definición exacta. Con el tiempo se han formulado muchas definiciones y los términos asociados, como Aprendizaje Automático y Aprendizaje Profundo (que se describirán más adelante en este capítulo y a fondo en el Capítulo 2) han ganado terreno, lo que contribuye a una mayor confusión.

El aspecto *artificial* de la IA es bastante claro: se refiere a todo lo que no es natural y, en este caso, hecho por el hombre. También se puede representar mediante el uso de términos como *máquinas*, *computadoras* o *sistemas*. El término *inteligencia* es un concepto mucho más controvertido, lo que explica por qué todavía no se ha llegado a un consenso sobre cómo definir la IA, ni siquiera entre los expertos (Miaihle y Hodes, 2017).

John McCarthy, a quien se le considera el padre de la IA, en 1956 definió la IA como "la ciencia e ingeniería para fabricar máquinas inteligentes".[4] Un enfoque importante para definir la IA, con base en un experimento ideado por Alan Turing, considera las similitudes entre las máquinas y los seres humanos en cuanto a la demostración de la inteligencia (véase el Recuadro 1.1). Esta prueba está bastante obsoleta en relación con los estándares tecnológicos actuales, sin embargo, es común que todavía se haga referencia a ella como uno de los primeros métodos para comprender la inteligencia de una máquina.

Recuadro 1.1: La "prueba de Turing"

En 1950, el matemático inglés Alan Turing desarrolló una prueba, que posteriormente recibió su nombre, la cual fue diseñada para determinar si una máquina (computadora) podía ser considerada inteligente. En la prueba se incluyó a tres participantes: un evaluador humano haría preguntas y un humano y una máquina escribirían las respuestas. La prueba define una máquina inteligente como una máquina que formula respuestas que el evaluador no puede distinguir de las del humano.

Fuente: https://searchenterpriseai.techtarget.com/definition/Turing-test.

Muchas definiciones generales reflejan este enfoque, incluyendo la definición de los sistemas de IA utilizada por la OCDE (Recuadro 1.2), que ha sido aceptada por 42 gobiernos nacionales. Además de ser máquinas que imitan a los humanos, la IA también puede entenderse como el área de conocimiento asociado con el diseño de estas máquinas o "la disciplina de crear algoritmos con la capacidad de aprender y razonar" (OCDE, 2018a).

Recuadro 1.2: Definición de la OCDE de los sistemas de IA

Un sistema basado en máquinas que puede, para un conjunto determinado de objetivos definidos por el ser humano, hacer predicciones, recomendaciones o tomar decisiones que influyen en entornos reales o virtuales. Los sistemas de IA están diseñados para funcionar con diversos niveles de autonomía. Además, la IA son "máquinas que realizan funciones cognitivas similares a las de los humanos".

Fuente: OECD (2019), Artificial Intelligence in Society, www.oecd.org/going-digital/artificialintelligence-in-society-eedfee77-en.htm; *OECD (2019), Recommendation of the Council on Artificial Intelligence,* https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[4] www.sciencedaily.com/terms/artificial_intelligence.htm.

Si bien la definición de la OCDE es aplicable en todos los sectores, otros expertos y organizaciones ven a la IA desde su propia perspectiva. Por ejemplo:

- El Grupo de expertos de alto nivel en Inteligencia Artificial (*High-Level Expert Group on Artificial Intelligence*) de la Comisión Europea (2019b) define la IA como "sistemas que muestran un comportamiento inteligente al analizar su entorno y tomar acciones, con cierto grado de autonomía, para lograr objetivos específicos".
- En el sector financiero, la autoridad regulatoria de Luxemburgo considera que las soluciones de IA son aquellas que "se centran en un número limitado de tareas inteligentes y se utilizan para ayudar a los seres humanos en el proceso de toma de decisiones" (CSSF, 2018).
- El Gobierno del Reino Unido define la IA como un "área de investigación que abarca la filosofía, la lógica, la estadística, la informática, las matemáticas, la neurociencia, la lingüística, la psicología cognitiva y la economía" que utiliza "tecnología digital para crear sistemas capaces de realizar tareas que comúnmente se cree que requieren inteligencia."[5]
- En lugar de articular su perspectiva sobre la "IA", la Asociación de Normas del IEEE se enfoca en los Sistemas Autónomos e Inteligentes (A/IS), que se centran más en los aspectos prácticos.[6]

Esta lista podría extenderse a lo largo de muchas páginas. Existen muchas definiciones o perspectivas para la IA debido, en gran parte, a que las nociones de lo que constituye la inteligencia son subjetivas. Por ejemplo, Howard Gardner (1983), en su libro *Estructuras de la mente: la teoría de las inteligencias múltiples*, propone una teoría en la que se reconocen ocho diferentes habilidades que conforman las diferentes partes de la inteligencia humana, incluyendo: habilidades del tipo rítmico-musical, lingüístico-verbal, lógico-matemático y habilidades interpersonales e intrapersonales. Desde esta perspectiva, la IA podría referirse a una máquina que es capaz de escribir música o resolver ecuaciones matemáticas tanto como a una máquina que puede expresar simpatía o amabilidad.

Otro factor que complica aún más las concepciones de "inteligencia" y lo que implica la IA, es el *tiempo*. Lo que pudiera considerarse inteligente puede evolucionar con el paso del tiempo. Muchas aplicaciones incorporadas extraordinarias disponibles en las computadoras o los teléfonos inteligentes, en un inicio se consideraron una forma de IA (por ejemplo, aplicaciones de mapas como Google Maps, que revisan cientos de miles de puntos de datos con el fin de proporcionar rutas óptimas), pero ahora se consideran características comunes. Asimismo, se suele suponer que los sistemas que realizan tareas al nivel de los estándares humanos o más allá de estos, son inteligentes cuando el factor subyacente en realidad es una potencia de procesamiento suficientemente grande. Tal y como lo señaló un colaborador de la consulta pública de este manual, "la velocidad puede hacerse pasar por inteligencia, pero no lo es."

Un ejemplo de estas percepciones cambiantes es el ajedrez, un juego que se asocia tradicionalmente a la IA. Hasta cierto punto, una computadora que puede jugar al ajedrez

[5] www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificialintelligence-in-the-public-sector.
[6] https://standards.ieee.org/industry-connections/ec/autonomous-systems.html.

contra un humano podría considerarse IA. Una vez que se enseñó a las computadoras a jugar el juego, el siguiente objetivo fue observar si una computadora inteligente podía vencer a un jugador humano y, posteriormente, vencer a los mejores jugadores humanos de ajedrez. Muchas tareas que las computadoras realizan y que en un momento dado podían considerarse IA ahora se describen con otros términos, como la automatización. Un artículo reciente, en forma de broma, hizo la siguiente distinción[7] "¿Cuál es la diferencia entre la IA y la automatización? Bueno, la automatización es lo que podemos hacer con las computadoras y la IA es lo que desearíamos poder hacer. En cuanto descubrimos cómo hacer algo, deja de ser IA y empieza a ser automatización." El término *Efecto IA* se ha acuñado para describir la naturaleza cambiante del término "IA" a medida que sus capacidades se normalizan (Recuadro 1.3).

Recuadro 1.3: El efecto IA

El fenómeno mediante el cual "los investigadores de la Inteligencia Artificial (IA) logran un hito que durante mucho tiempo se pensó que significaba el logro de la verdadera inteligencia artificial, por ejemplo, ganarle a un humano en el ajedrez, y de repente se degrada a IA *no verdadera*."

Fuente: https://medium.com/@katherinebailey/reframing-the-ai-effect-c445f87ea98b.

Aunque el debate sobre la definición de la IA es fascinante, el objeto de este manual es proporcionar una visión general de la IA, y analizar las posibles aplicaciones y consideraciones relativas a la innovación en el sector público y su transformación. El primer paso para comprender y determinar el posible impacto de la IA en el sector público es explorar cuán inteligentes son las máquinas hoy en día.

IA general frente a IA estrecha

A pesar de la falta de consenso para definir la IA, en general los expertos reconocen dos amplias perspectivas que ayudan a establecer las expectativas sobre cuán "inteligente" puede ser la IA. La primera es la **IA general**, también denominada "IA fuerte" o la perspectiva de la "Inteligencia Artificial General" (IAG). La segunda es la **IA estrecha**, también denominada "IA débil", "IA aplicada" o "Inteligencia Artificial Estrecha" (IAE).

IA general: lograr cualidades similares a las de los humanos

La IA general se refiere a la idea de que la inteligencia humana general, que abarca diferentes áreas y capacidades, podría ser igualada o incluso superada por las máquinas. Algunas investigaciones utilizan el término "Súper Inteligencia Artificial", o "superinteligencia", para referirse a sistemas hipotéticos de IA general que podrían superar por mucho las capacidades de los humanos (Bostrom, 2014).

[7] https://arstechnica.com/features/2019/04/from-ml-to-gan-to-hal-a-peak-behind-the-modern-artificialintelligence-curtain.

Un supuesto riesgo o dilema que plantea la IA general, que con frecuencia exageran los medios de comunicación populares, es la creencia de que los intereses de dicho sistema de IA pudieran no estar de acuerdo necesariamente con los de la humanidad y que, en última instancia, pudiera desafiar a los seres humanos. Dichas representaciones de la IA en los medios de comunicación masiva y la cultura popular han captado la imaginación del público y pueden contribuir a la desconfianza en la tecnología impulsada por la IA, aunque existen opiniones encontradas entre los investigadores de la IA en cuanto a la posibilidad de tal escenario.

Algunos acontecimientos recientes de la investigación en las neurociencias permiten comprender mejor cómo funciona nuestro cerebro y parecen indicar que la IA general puede ser posible. De hecho, existen razones para creer que diferentes funciones de los cerebros artificiales podrían combinarse para lograr funciones cognitivas más complejas (Goodfellow, Bengio y Courville, 2016). Sin embargo, aún falta mucho por comprender sobre el cerebro humano, y por lo tanto, sobre los artificiales. Si bien la ciencia actual deja la puerta abierta, la IA general, por el momento, permanece en el mundo de ciencia ficción. Las opiniones de los expertos están divididas en cuanto a la posibilidad de volver la IA general una realidad y el tiempo que tiene que pasar para ello. Estudios recientes muestran que muchos investigadores de la IA creen que existe una probabilidad del 50% de que la IA sea capaz de superar a los humanos en todas las tareas dentro de unos 45 años, mientras que otros piensan que esto nunca sucederá (Grace et al., 2018; Muller y Bostrom, 2014). Mientras tanto, la IA estrecha, tal y como se explica a continuación, ya proporciona muchos beneficios tangibles para la sociedad, aunque esto también tiene sus propias limitaciones y desafíos que hay que tener en cuenta.

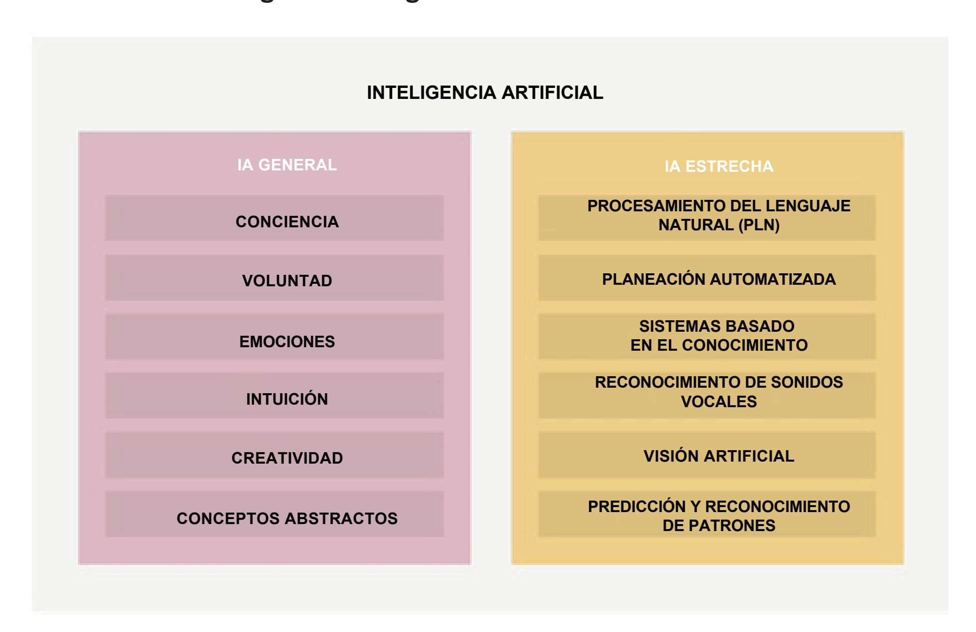
IA estrecha: una visión más detallada de la IA

En la actualidad, toda la IA es IA estrecha. Ningún algoritmo, máquina o computadora de IA es capaz de superar a los humanos en una gama de tareas. La perspectiva de la IA estrecha, a diferencia de la IA general, se preocupa menos por la creación de una superinteligencia unificada, y en lugar de ello acepta y aprovecha la noción de que los humanos y las computadoras tienen diferentes fortalezas y competencias relativas. La IA estrecha aprovecha el hecho de que las computadoras se destacan por su desempeño cuando se trata de procesar grandes cantidades de datos de manera rápida y coherente (véase el Capítulo 2 para conocer más sobre datos), así como para ejecutar tareas basadas en reglas lógicas y explícitas, mientras que los humanos son aún más eficientes en el manejo de situaciones ambiguas o aquellas que requieren intuición, creatividad, emoción, juicio y empatía.

La IA estrecha también refleja las capacidades actuales de las máquinas que permiten a las computadoras realizar tareas de manera inteligente en áreas específicas, pero que no permiten que estas áreas se unan para producir una inteligencia más completa. A este respecto, con el paso de los años diferentes comunidades de investigadores han respaldado el desarrollo de varias ramas de la IA, cada una de ellas enfocada en un conjunto específico de tareas que, en general, están en consonancia con diferentes capacidades humanas (Frank et al., 2019). Si bien existen diversas ramas dentro del área mucho más amplia de la IA, algunas de las áreas de mayor aplicación y publicidad son las siguientes:

- **Visión artificial** se refiere a la capacidad de la IA para procesar y sintetizar datos visuales (por ejemplo, detectar y clasificar objetos con base en las imágenes o videos) y realizar tareas como el reconocimiento facial y la interpretación de situaciones. Además de procesar los datos visuales existentes, algunas aplicaciones implican la construcción de datos visuales, como la creación de modelos en 3D a partir de imágenes en 2D (Soltani et al., 2017).
- **Procesamiento del lenguaje natural (PLN)** se refiere a la capacidad que tienen las computadoras para manipular e interpretar el lenguaje humano y realizar diversas tareas como la traducción o el análisis de textos. Además de procesar el lenguaje existente, las aplicaciones de PLN también pueden generar un nuevo lenguaje hablado o escrito.
- **Reconocimiento de sonidos vocales** se refiere a la capacidad que tienen las computadoras para analizar archivos de audio con el fin de reconocer e interpretar el lenguaje hablado.
- **Sistemas basados en el conocimiento** registran y almacenan hechos en una "base de conocimientos", y posteriormente utilizan un "motor de inferencias" para inferir información de la base de conocimiento con el fin de resolver problemas, con frecuencia a través de reglas programadas del tipo SI…ENTONCES. [8] Por ejemplo, los "sistemas expertos" codifican los conocimientos de los expertos para ayudar a resolver problemas a través de reglas del tipo SI…ENTONCES.
- **Planeación automatizada** se refiere a la capacidad que tienen las máquinas para diseñar de manera automática y autónoma medidas o estrategias para lograr un objetivo, incluyendo la anticipación de los efectos de diferentes enfoques.[9]

Figura 1.2: IA general frente a IA estrecha



Fuente: OPSI.

8 https://searchcio.techtarget.com/definition/knowledge-based-systems-KBS.
9 https://github.com/pellierd/pddl4j/wiki/Automated-planning-in-a-nutshell.

En el Capítulo 2 se analizan con más detalle algunos de estos enfoques. La lista anterior no es en absoluto exhaustiva, ya que existen muchas ramas de la IA con la capacidad de realizar diferentes tareas. Además, estas ramas diversas y sus comunidades asociadas no se excluyen mutuamente, sino que evolucionan de manera que a veces se conectan o se superponen. Por ejemplo, los avances en los ámbitos del reconocimiento de sonidos vocales y el PLN pueden influir y beneficiarse mutuamente ya que ambos se relacionan, en cierta medida, con el análisis del lenguaje. La visión artificial en combinación con el PNL[10] promete resultados en la descripción y el subtitulado de imágenes o videos.

Aumento de la Inteligencia Artificial

La naturaleza de los progresos que se están realizando en la IA estrecha demuestra que los seres humanos y las máquinas, en lugar de competir, ambos tienen puntos fuertes y débiles, por lo que podrían beneficiarse de la colaboración mutua. Al interactuar entre sí, los humanos y las máquinas pueden resolver problemas y lograr mejores resultados de los que podría lograr cada uno por su cuenta. Dicho enfoque, que hace hincapié en las interacciones entre los seres humanos y las máquinas, puede denominarse Aumento de la Inteligencia Artificial o simplemente Aumento de la Inteligencia (Carter y Nielsen, 2017).[11] En sentido más figurado, el término "centauro" se ha utilizado para describir la colaboración entre el hombre y la inteligencia artificial (haciendo referencia a la criatura mitológica mitad humana, mitad caballo) (Case, 2018).

De hecho, hay muchos ejemplos en los que equipos de humanos y computadoras que trabajan juntos han sido capaces de vencer no sólo a equipos de humanos, sino también a equipos de computadoras. En el Recuadro 1.4 se incluyen algunos ejemplos.

Recuadro 1.4: Aumento de la inteligencia: Ejemplos de colaboración entre seres humanos y máquinas

Un enfoque importante en la cooperación entre humanos y la IA, especialmente en tareas creativas, es el uso de algoritmos evolutivos y genéticos. Esto suele implicar la definición de un problema, seguida de la generación de una serie de posibles soluciones. Aunque no todos estos procesos algorítmicos implican interacción humana, pueden diseñarse para garantizar que una persona se encargue de seleccionar entre las posibles soluciones que se ofrecen.

Por ejemplo, el profesor de informática Sung-Bae Cho creó una herramienta de IA que genera diferentes diseños de atuendos. El usuario selecciona los diseños que desea conservar, y estas opciones se vuelven a introducir en la IA, permitiéndole aprender y

[10] www.researchgate.net/publication/311625991_Computer_Vision_and_Natural_Language_Processing_Recent_Approaches_in_Multimedia_and_Robotics.

[11] En la investigación y comercialización de la IA ha evolucionado un amplio espectro de interacción hombre-máquina. Por ejemplo, la "inteligencia asistida" se refiere a las situaciones en las que "la IA ha sustituido a muchas de las tareas repetitivas y estandarizadas que realizan los seres humanos"; la "inteligencia aumentada" se refiere a los casos en que "los seres humanos y las máquinas aprenden entre sí y redefinen el alcance y la profundidad de lo que hacen juntos", y la "inteligencia autónoma" describe las situaciones en que "los sistemas adaptables/continuos se hacen cargo en algunos casos". Para obtener más información véase www.recode.net/sponsored/11895802/what-artificial-intelligence-really-means-to-business y https://medium.com/cognilytica/assisted-intelligence-vs-augmented-intelligence-1db96ef3457f.

Recuadro 1.4: Aumento de la inteligencia: Ejemplos de colaboración entre seres humanos y máquinas *(Cont.)*

evolucionar para que pueda generar nuevos diseños. El proceso se repite hasta que la IA produce resultados que el usuario considera satisfactorios.

Se han desarrollado sistemas similares en áreas tan diversas como la ingeniería industrial, la medicina y los videojuegos. En el caso del diseño industrial, los ingenieros pueden configurar restricciones y generar planos de edificios o sistemas mecánicos y seleccionar los que mejor se adapten. En medicina, los investigadores pueden producir nuevos medicamentos utilizando algoritmos evolutivos para generar combinaciones de moléculas y, posteriormente, eliminar las que no producen beneficios para la salud. En la industria de los videojuegos, los mismos principios permiten que los diseñadores generen rápidamente objetos como edificios, calles y autos que, de otro modo, tardarían varias horas.

A fin de ayudar en estas colaboraciones entre seres humanos y máquinas, las soluciones conocidas como "interfaces generativas" permiten a los usuarios humanos interactuar y comprender todas las posibles soluciones generadas.

Esta nueva forma de cooperación y colaboración permite que se creen rápidamente soluciones nuevas, interesantes y creativas para los problemas del mundo real. No sustituye la necesidad de la creatividad humana, sino que se centra en aumentar las capacidades creativas permitiendo que los seres humanos interactúen y exploren miles de soluciones posiblemente únicas de una manera rápida y eficiente.

Fuente: https://distill.pub/2017/aia; https://jods.mitpress.mit.edu/pub/issue3-case; Cho, S-B. (2002), "Towards creative evolutionary systems with interactive genetic algorithm", *Applied Intelligence*, Vol. 16/2, pp. 129-138, https://link.springer.com/article/10.1023%2FA%3A1013614519179; https://medium.com/generativedesign/evolving-design-b0941a17b759; https://becominghuman.ai/understanding-evolutionary-algorithms58f7a2845537; http://dougengelbart.org/content/view/138.

En el caso de los enfoques de Aumento de la Inteligencia, es importante desarrollar un diseño adecuado para la interacción entre el ser humano y la computadora. Este factor puede ser tan importante como la potencia bruta de procesamiento cuando se considera el rendimiento general (Case, 2018). Un operador humano que trabaja con una interfaz mal diseñada conectada a dos supercomputadoras puede ser menos eficaz que un operador humano que trabaja con una sola computadora regular bien diseñada.

En cuanto al uso del Aumento de la Inteligencia en el sector público, se podría invertir en desarrollar el uso de la inteligencia artificial para incrementar las capacidades de los funcionarios públicos y que de esta manera éstos puedan hacer mejor su trabajo. Los funcionarios públicos podrían procesar y consultar grandes cantidades de datos con mayor rapidez, lo que les permitiría evaluar opciones alternativas o más personalizadas al prestar servicios a los ciudadanos.

En conclusión, la distinción entre la IA general y la IA estrecha es significativa y refleja perspectivas diferentes pero compatibles en torno al potencial de la IA. Su situación actual, hablando de la IA estrecha, ya plantea oportunidades y desafíos apremiantes que la sociedad y los gobiernos deben abordar. Al mismo tiempo, vale la pena crear estrategias y estudiar los posibles riesgos de la IA general en caso de que algún día surjan. De acuerdo con lo que se

establece en la Ley de Amara, “tendemos a sobrestimar los efectos de una nueva tecnología a corto plazo, mientras que subestimamos su efecto a largo plazo”[12]

Independientemente de las perspectivas, es posible sobrestimar o subestimar los logros y los posibles impactos de la IA. La siguiente sección echa un vistazo a la historia del desarrollo de la IA y trata de entender la dinámica que sustenta el entusiasmo actual. Si bien la evolución a largo plazo de la IA es en esencia incierta, los gobiernos deberían tratar de explorar las oportunidades inmediatas que ofrece la IA en la actualidad, y al mismo tiempo establecer marcos para prepararse para los cambios tecnológicos y científicos a largo plazo.

Entusiasmo renovado por la IA

Estaciones de la IA

Como se mencionó anteriormente, la IA no es un concepto nuevo. El tema ha estado y pasado de moda con tanta frecuencia que las comunidades de la IA hablan metafóricamente de los “inviernos” de la IA para referirse a las épocas en que la gente le da la espalda a la IA después de que no ha estado a la altura de sus expectativas.

Los primeros avances científicos generaron entusiasmo en torno a los posibles logros de la IA. Las personas trazaron grandes expectativas y sobrestimaron lo que se podía lograr en ese momento. Por ejemplo, en la década de 1950, cuando se comenzó a estudiar la IA, acontecimientos recientes hacían pensar que los humanos estaban a punto de crear máquinas autodidactas y que podían utilizarse para la automatización de varias tareas, incluyendo la traducción y la toma de decisiones, pero éstas no estuvieron a la altura de las expectativas. En la década de 1980, la introducción de los enfoques de IA basados en reglas (del tipo SI…ENTONCES) (los cuales se describen en el Capítulo 2) también provocó entusiasmo; sin embargo, la emoción disminuyó rápidamente cuando los sistemas basados en reglas resultaron difíciles de implementar a gran escala.

Si las expectativas que se fijan son demasiado altas y la IA es incapaz de cumplir sus promesas, las personas centran su atención en otras tecnologías o áreas. La financiación y las inversiones en investigación, que son cruciales para que IA siga desarrollándose, se reorientan hacia otras áreas que prometen mejores resultados a corto plazo. Impulsada por los avances tecnológicos en las técnicas de Aprendizaje Automático, la IA ha experimentado un entusiasmo renovado en los últimos años.

Generadores de la oleada de entusiasmo actual

Toda la cobertura e interés que han rodeado a la IA en los últimos años, tanto en el sector privado como en el público, han llevado a muchos expertos a referirse a la agitación actual como una “primavera de la IA”. Se pueden mencionar muchos factores que explican el continuo y creciente optimismo, tal y como se describe a continuación.

Madurez del área

Se ha acumulado un gran conocimiento a partir de diversos proyectos lanzados en las últimas décadas. Se han perfeccionado viejos algoritmos y modelos, y han surgido otros nuevos. Se han desarrollado y perfeccionado los lenguajes y marcos de programación y se

12 https://spotlessdata.com/blog/amaras-law.

han creado muchas aplicaciones nuevas a medida que más personas se familiarizan con la IA. Por ejemplo, la idea de las neuronas artificiales ha existido desde los años 40; sin embargo, el desarrollo de la IA de Aprendizaje Profundo sólo tomo impulso durante la última década (para conocer más detalles sobre el Aprendizaje Profundo véase el Capítulo 2)[13]

Mejor tecnología

Las computadoras en la actualidad son más económicas, tienen más potencia de procesamiento y requieren mucho menos espacio físico. Este aumento de la potencia de procesamiento permite que los dispositivos ejecuten programas más grandes y complejos, y que procesen más datos con mayor rapidez. Los costos de almacenamiento de datos también han disminuido drásticamente; el costo de un gigabyte de almacenamiento cayó de USD 1 millón en 1967 a aproximadamente 2 centavos en 2017[14]

Democratización de las computadoras y la programación

Si bien la tecnología ha mejorado, también es cierto que cada vez está al alcance de un mayor número de personas. Los nuevos usuarios de hoy en día también están más conectados y mejor equipados para aprender e intercambiar información sobre la IA. Las plataformas y herramientas de colaboración respaldadas por comunidades dinámicas están haciendo posible la programación y la codificación no sólo para los expertos y las empresas, sino también para personas de todos los orígenes. Por ejemplo, GitHub y Kaggle permiten que las personas colaboren en soluciones digitales (véanse los Recuadros 1.5 y 1.6). Esto también permite que las ideas y soluciones de enfoques participativos surjan de maneras que no eran a menudo posibles en el pasado. Los cursos y tutoriales que se proporcionan de forma gratuita en línea, incluyendo los que ofrece el sector público, también contribuyen a esta democratización (véanse los recuadros 1.7 y 1.8).

Recuadro 1.5: GitHub en colaboración con la IA

GitHub es una plataforma web popular que sirve como depósito de códigos y como red social. Los usuarios pueden alojar y compartir públicamente códigos de la computadora en los depósitos de códigos. A menudo, se otorga la licencia del código como software libre y de código abierto (FOSS, por sus siglas en inglés), lo que permite que otros puedan descargarlo sin costo y aplicar el código fuente a su propio trabajo y contribuir en los proyectos de otros usuarios. La plataforma también permite el desarrollo en colaboración, en el que varios usuarios realizan cambios para mejorar el código, los cuales se rastrean a través de Git, un sistema de control de versiones. Por ejemplo, en 2017 el Gobierno de los Estados Unidos llevó a cabo un Proyecto piloto de asistente personal de IA a nivel federal (*Federal AI Personal Assistant Pilot*) para la introducción efectiva, eficiente y responsable de los asistentes personales inteligentes disponibles (por ejemplo, Alexa de Amazon, el Asistente de Google, Cortana de Microsoft, los bots conversacionales de Facebook Messenger) en los programas gubernamentales. GitHub se utilizó como una herramienta de convocatoria y espacio de colaboración para el proyecto piloto.

Fuente: OPSI; https://github.com/GSA/AI-Assistant-Pilot.

13 https://towardsdatascience.com/rosenblatts-perceptron-the-very-first-neural-network-37a3ec09038a.
14 www.computerworld.com/article/3182207/cw50-data-storage-goes-from-1m-to-2-cents-pergigabyte.html.

La plataforma Kaggle mezcla el espíritu competitivo con la colaboración para producir rápidamente soluciones a problemas específicos.

Recuadro 1.6: Uso de Kaggle para enfrentar los desafíos de la IA

Kaggle es una comunidad en línea de competencias en ciencias de datos, incluyendo aquellas relacionadas con la IA. La plataforma en sí es propiedad de Google y permite a los usuarios alojar y publicar conjuntos de datos. Posteriormente, los editores pueden crear desafíos basados en dichos conjuntos de datos proporcionando una descripción del problema que tratan de resolver. Los científicos de datos pueden participar en la competencia de manera individual o en equipo proponiendo diferentes modelos que están disponibles al público y se califican con base en los criterios de evaluación especificados por el anfitrión antes de la competencia. Las soluciones con mayor puntuación suelen recibir premios monetarios del anfitrión, y a menudo las soluciones se ofrecen como software de código abierto para que cualquiera pueda utilizarlas. La plataforma ofrece funciones comunitarias para debatir el problema e intercambiar ideas sobre los retos relativos a los conjuntos de datos en un espíritu de colaboración para la resolución de problemas.

Detección de la neumonía por medio de la IA

En agosto de 2018, la Sociedad Radiológica de Norteamérica (RSNA) se asoció con organizaciones como los Institutos Nacionales de Salud de los Estados Unidos para organizar un concurso a través de Kaggle para desarrollar un sistema de detección automática de casos de neumonía con una aplicación de IA basada en radiografías de tórax. Se ofreció un premio de USD 30,000. El concurso se desarrolló hasta finales de octubre de 2018 y participaron más de 1,400 equipos. Finalmente, la RSNA otorgó un reconocimiento a diez equipos durante su reunión anual en noviembre de 2018. En particular, el equipo con mayor puntaje, el de Ian Pan, estudiante de medicina en ese momento, y Alexandre Cadrin-Chênevert, radiólogo e ingeniero informático, desarrollaron una combinación de modelos de aprendizaje profundo que logró los mejores resultados para detectar casos de neumonía y que podría tener efectos significativos para el tratamiento de esta enfermedad. El código de esta solución está disponible en GitHub.

Fuente: www.github.com/i-pan/kaggle-rsna18; www.kaggle.com/c/rsna-pneumonia-detection-challenge#Prizes; www.kaggle.com/c/rsna-pneumonia-detection-challenge/discussion/70421; www.rsna.org/en/education/ai-resources-and-training/ai-image-challenge/RSNA-Pneumonia-Detection-Challenge-2018.

Los cursos en línea masivos y abiertos (MOOC), los tutoriales y varios sitios web participan en la democratización del conocimiento, y en particular, de la codificación, por lo general sin ningún costo (Recuadro 1.7).

Recuadro 1.7: Cursos gratuitos útiles sobre IA

Existen muchos cursos disponibles sobre IA. A continuación, se presenta una selección de cursos gratuitos útiles para las personas que trabajan en el sector público:

- Elementos de la IA (*Elements of AI*) es un curso gratuito en línea de seis partes sobre IA desarrollado en conjunto con la Universidad de Helsinki y Reaktor, una organización de servicios de consultoría y agencias. Este curso sirve como una introducción a la IA para los no expertos. http://course.elementsofai.com.
- Introducción a la Inteligencia Artificial (*Introduction to Artificial Intelligence*) es otro curso gratuito en línea que ofrece Udacity. https://eu.udacity.com/course/intro-to-artificial-intelligence-cs271.
- *fast.ai* ofrece cursos de aprendizaje profundo, aprendizaje automático y procesamiento del lenguaje natural, entre otros. Los cursos por lo general están enfocados en los participantes que ya cuentan con experiencia en materia de codificación. www.fast.ai.
- Las plataformas digitales Coursera y edX ofrecen acceso gratuito a cursos en línea que se dirigen a un público más avanzado. Algunos cursos ofrecen certificación, la cual puede tener un costo. Dos cursos muy conocidos son Aprendizaje Automático (*Machine Learning*) (www.coursera.org/learn/machine-learning) e IA para todos (*AI for Everyone*) (www.coursera.org/learn/ai-for-everyone). Otros se pueden encontrar en www.coursera.org/courses?query=artificial%20intelligence y www.edx.org/course?search_query=artificial+intelligence.
- Google ofrece cursos gratuitos sobre varios temas de IA https://ai.google/education.
- En los Países Bajos, la fundación *IA para el bien social* (*AI for GOOD*), el Centro de conocimientos del ICAI (*ICAI Knowledge Centre*) y la empresa Elephant Road colaboraron para crear el Curso nacional de IA (*De National AI-Cursus*). Este curso gratuito sobre los elementos esenciales de la IA se ofrece a los holandeses en su propio idioma. También se ofrece una versión en inglés, la cual se agregó recientemente. Véase https://app.ai-cursus.nl.

Los gobiernos de algunos países imparten capacitación a ciudadanos de todas las edades y a funcionarios públicos sobre los conocimientos básicos de cómo utilizar una computadora, navegar por Internet y buscar información. Cada vez más, algunos también proporcionan conocimientos especializados sobre IA (véase el Recuadro 1.8). Estas iniciativas aumentan los conocimientos informáticos y estimulan la capacidad de las personas para aprovechar el trabajo existente sobre IA para su uso personal y profesional.

Recuadro 1.8: Iniciativas de aprendizaje impulsadas por el gobierno

En 2018, el Gobierno de Francia aprobó una nueva Estrategia nacional contra la brecha digital (*Stratégie nationale pour un Numérique inclusif*), que busca brindar apoyo a los ciudadanos y residentes de Francia para que adquieran conocimientos, habilidades y autonomía en materia de tecnología digital. Como parte de esta estrategia, el gobierno lanzó el Pase digital (*Pass numérique*), que ofrece capacitación gratuita en tecnologías digitales.

Los gobiernos de todo el mundo también están empezando a adoptar medidas para incrementar las habilidades de sus funcionarios públicos. Por ejemplo, el Instituto francés de Gestión Pública y Desarrollo Económico (IGPDE, por sus siglas en francés), que forma parte del Ministerio de Economía y Finanzas de Francia, ofrece muchos cursos de capacitación diferentes, incluyendo sesiones breves de un día, como “Transformación digital del estado y los datos” (*Digital transformation of the state and data*) e “Inteligencia artificial, ciencia de los datos: Nuevos desafíos económicos” (*Artificial intelligence, data science: New economic challenges*). Su objetivo es dotar a los funcionarios públicos de conocimientos básicos sobre IA y sus oportunidades y desafíos.

El Gobierno de Singapur ofrece talleres de IA abiertos a funcionarios públicos y, en particular, a los gerentes de nivel medio y superior. Su objetivo es aumentar la alfabetización digital y proporcionar conocimientos fundamentales sobre el potencial de la IA para el trabajo público y las organizaciones públicas. (Véase el Recuadro 4.22 del Capítulo 4 para obtener más detalles sobre la Academia Digital de la Escuela de la Función Pública de Canadá [*Canada School of Public Service's Digital Academy*] y la forma en que proporcionan recursos a los funcionarios públicos en materia de inteligencia artificial)

Fuente: https://societenumerique.gouv.fr/inclusion; www.cscollege.gov.sg/programmes/pages/display%20programme.aspx?epid=cn5g9p9ecwsdnnrg2eu7pshwu1; www11.minefi.gouv.fr/catalogue-igpde/2019/co/7783.html; www11.minefi.gouv.fr/catalogue-igpde/2019/co/8618.html; www.leparisien.fr/yvelines-78/davantage-d-informatique-a-la-maison-de-la-reussite-08-07-1998-2000150304.php; www.netpublic.fr/net-public/espaces-publics-numeriques/presentation.

Disponibilidad de datos y aprendizaje automático

A menudo, la abundancia de datos se cita como el principal impulsor del entusiasmo actual en torno a la IA. Por ejemplo, un factor clave que contribuye al auge de las aplicaciones de IA es el hecho de que muchas interacciones diarias son ahora digitales o asistidas digitalmente y generan un volumen importante de datos. Se estima que el 90% de los datos del mundo se crearon sólo en los últimos años y que las tasas de generación de datos siguen acelerándose (Marr, 2018). Este fenómeno se denomina “Big Data” (OCDE, 2015a) y se caracteriza por

- *Velocidad*. Los datos se generan y procesan a una velocidad más rápida que en cualquier otro momento de la historia.
- *Volumen*. En la actualidad existe una inmensa cantidad de datos generados y almacenados.
- *Variedad*. Los datos se presentan en diferentes formas y formatos, incluyendo texto, imágenes, video y audio.

Los organismos del sector público ocupan una posición interesante en lo que respecta a la generación de datos. Muchas operaciones gubernamentales se relacionan con el mantenimiento de los registros civiles (por ejemplo, nacimientos, matrimonios). Los gobiernos también conservan una gran cantidad de datos de otro tipo, incluyendo datos geoespaciales y meteorológicos procedentes de satélites, registros de bienes y registros de salud y seguridad, entre muchos otros.

En los últimos años, los gobiernos han buscado publicar datos gubernamentales en formatos legibles por máquina a través de políticas de datos abiertos gubernamentales (DAG) y portales asociados para conjuntos de datos e Interfaces de Programación de Aplicaciones.[15,16] Lo anterior permite que los datos estén disponibles para que los sistemas de IA puedan aprovecharlos.

En el sector privado, las empresas recopilan datos sobre sus clientes, empleados y proveedores con el fin de administrar sus negocios de manera más eficiente. Los usuarios de los servicios basados en la web también generan diversos tipos de datos. Por ejemplo, el uso de teléfonos móviles y computadoras, la navegación por la Web, el uso de redes sociales y el realizar compras generan una cantidad importante de datos, contribuyendo así al fenómeno del Big Data.

Con la llegada del Internet de las cosas (IoT)[17] y el desarrollo de la Industria 4.0 (OCDE, 2017c),[18] se dispone de importantes fuentes adicionales de datos, incluyendo el acceso a los datos generados por todo tipo de dispositivos electrónicos, sensores, aparatos, máquinas y vehículos.

La enorme y creciente cantidad de datos disponibles, junto con las tecnologías de procesamiento necesarias, han impulsado el avance del Aprendizaje Automático y el Aprendizaje Profundo, subconjuntos de la IA que representan un cambio de paradigma en la forma de utilizar las computadoras. En lugar de que los humanos programen de forma manual las máquinas con reglas sobre cómo pensar, la enorme cantidad de datos de los que se dispone puede alimentar a las computadoras para que puedan aprender las reglas por sí mismas y producir nuevos conocimientos (véase la Figura 1.3 y 1.4). En el Capítulo 2 se proporcionan más detalles sobre el Aprendizaje Automático y otros enfoques de la IA.

---

15 La OCDE ha publicado varios informes y otros productos relacionados con el estado actual de las estrategias e iniciativas de los DAG, así como los desafíos asociados, las lecciones aprendidas, los casos de estudio y las recomendaciones. Para obtener más información véase www.oecd.org/gov/digital-government/open-government-data.htm

16 Una API consiste en un conjunto de definiciones, protocolos y herramientas para crear software de aplicación. En pocas palabras, se trata de un conjunto de métodos que permiten que los componentes de software interactúen entre sí, posibilitando, por ejemplo, que los usuarios copien y peguen texto u otros tipos de datos de una aplicación a otra. Existen muchos tipos diferentes de API para sistemas operativos, aplicaciones o sitios web. Una buena API facilita el desarrollo de un programa, ya que proporciona todos los bloques de información, los cuales posteriormente son ensamblados por un programador. Para obtener más información véase www.webopedia.com/TERM/A/API.html.

17 El IoT incluye todos los dispositivos y objetos cuyo estado puede modificarse a través de Internet, con o sin la participación activa de los individuos. Ello incluye computadoras portátiles, enrutadores, sensores, servidores, tabletas y teléfonos inteligentes (OCDE, 2018b).

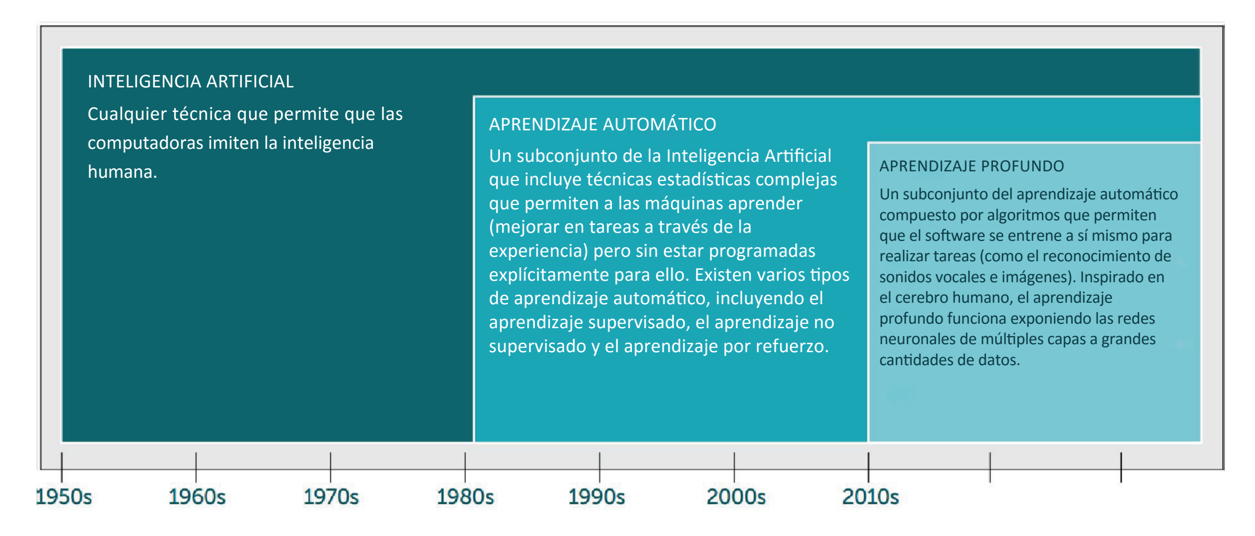
18 "Industria 4.0" o "la cuarta revolución industrial" son términos que se emplean para describir el surgimiento actual de una nueva revolución en la producción derivada de una confluencia de tecnologías. Esto abarca desde diversas tecnologías digitales (por ejemplo, la impresión en 3D, el Internet de las cosas, la robótica avanzada) y nuevos materiales (por ejemplo, de tipo bio o nano) hasta nuevos procesos (por ejemplo, la producción basada en datos, la Inteligencia Artificial, biología sintética).

Figura 1.3: Diferencia entre la programación clásica y el Aprendizaje Automático

reglas
datos
PROGRAMACIÓN CLÁSICA
respuestas
datos
respuestas
APRENDIZAJE AUTOMÁTICO
reglas

Fuente: European Commission (2018a), *Artificial Intelligence: A European Perspective*, https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-european-perspective.

Figura 1.4: El posicionamiento de la IA, el Aprendizaje Automático y el Aprendizaje Profundo



Fuente: https://static1.squarespace.com/static/5b156e3bf2e6b10bb0788609/t/5d2c43ca74551c000190105f/1563182032127/AI+and+international+Development_FNL.pdf.

En la siguiente sección se analizan las implicaciones de la IA en el sector público y de qué manera los propios gobiernos pueden utilizarla para mejorar su funcionamiento y cumplir sus misiones y objetivos a fin de servir a sus ciudadanos, residentes, empresas y otras organizaciones.

La IA y el sector público

En las secciones anteriores se ha definido a la IA en términos generales, pero ¿qué significa la IA en el sector público? ¿Qué importancia tiene para los dirigentes, administradores y otros funcionarios públicos? El OPSI ha observado, y las investigaciones lo han comprobado, que el sector público está utilizando cada vez más la IA. Sin embargo, la gran mayoría de las

investigaciones se centran en aspectos técnicos específicos o en el uso que el sector privado hace de la misma. De hecho, en una revisión bibliográfica reciente de casi 1700 estudios relativos a la IA, se encontró que solo 59 (3.5%) se centraban en el uso de la IA en el sector público (Gomes de Sousa et al., 2019).

A pesar de la falta de estudios, la IA es un tema importante y cada vez más recurrente en los gobiernos. Uno de los principales objetivos del sector público es formular y mejorar las leyes y políticas, suministrar bienes y servicios públicos a los ciudadanos y residentes, y proporcionar y mantener los instrumentos, recursos y estructuras necesarios para que los funcionarios públicos puedan desempeñar sus funciones. De acuerdo con este fin, los gobiernos desempeñan una serie de funciones en relación con la IA (véase el Recuadro 1.9). Si bien estas funciones son generales, este manual explora la forma en que se relacionan con la innovación y la transformación de los procesos, prácticas, políticas y servicios del sector público, en lugar de cómo afectan a la economía en general.

Recuadro 1.9: Funciones de la IA seleccionadas por el gobierno

En los estudios de la OCDE se han identificado una serie de funciones que los gobiernos pueden desempeñar en relación con la IA, a menudo de forma simultánea:

- **El gobierno como financista o inversionista directo.** Los gobiernos pueden asignar fondos para apoyar el desarrollo y la adopción de nuevas tecnologías. Algunos están trabajando de forma activa en planes de financiación relacionados con convocatorias de proyectos o licitaciones piloto. Dichos planes incluyen proyectos dentro del sector público, así como proyectos de I+D del sector privado cuyos resultados pueden aplicarse a toda la economía.
- **El gobierno como comprador inteligente y codesarrollador.** Los gobiernos pueden actuar como compradores inteligentes de las soluciones existentes mediante prácticas de adquisición innovadoras, o como codesarrolladores mediante asociaciones público-privadas (PPP, por sus siglas en inglés) y otras formas de colaboración a fin de crear soluciones nuevas o personalizadas. Los gobiernos pueden impulsar la innovación desde la perspectiva de la demanda dirigiendo el desarrollo de nuevas soluciones directamente hacia sus necesidades.
- **El gobierno como regulador o encargado de dictar normas.** Los ciclos de innovación acelerados de las tecnologías digitales emergentes exigen replantear los tipos de instrumentos normativos y reglamentarios utilizados y su implementación. Los gobiernos, en su papel de facilitadores y usuarios de las nuevas tecnologías digitales se enfrentan al desafío de determinar cómo, y en qué medida, regularlas para maximizar su potencial innovador minimizando los riesgos para los usuarios finales. Un aspecto de esta función consiste en evaluar el cumplimiento de los reglamentos y las normas y adoptar medidas necesarias cuando se infrinjan, según sea pertinente.
- **El gobierno como conciliador y organismo normativo.** Los gobiernos suelen tener la capacidad de reunir a las partes interesadas de diversos sectores del ecosistema de la IA (por ejemplo, ciudadanos y residentes, empresas, organizaciones e

Recuadro 1.9: Funciones de la IA seleccionadas por el gobierno *(Cont.)*

intelectuales) para ayudarles a alcanzar sus objetivos y comprender los múltiples aspectos de los problemas en cuestión. Los gobiernos también pueden ayudar a elaborar e implementar normas y normas extraoficiales en materia de tecnología en colaboración con dichas partes interesadas.

- **El gobierno como administrador de datos**. Los gobiernos son dueños, o tienen en su posesión en nombre de su pueblo, grandes cantidades de datos. Dichos datos pueden alimentar las tecnologías basadas en la IA, en especial cuando están bien administrados (en el Capítulo 2 véase el apartado Datos: el alimento de la IA).
- **El gobierno como usuario y suministrador de servicios.** Los gobiernos suministran servicios y herramientas habilitados o posibilitados por las tecnologías de IA, tanto al público en general como a las funciones administrativas. Por lo tanto, desempeñan una función en el uso y la adopción de las tecnologías por sí mismos.

Fuente: Ubaldi, B. et al. (2019), "State of the art in the use of emerging technologies in the public sector", *OECD Working Papers on Public Governance*, No. 31, OECD Publishing, París, https://doi.org/10.1787/932780bc-en; OPSI (los últimos tres puntos).

Dado que sus funciones difieren de las de otros sectores, algunos casos de uso y consideraciones respecto a la aplicación de la IA en el sector público pueden ser más relevantes que otros. Por ejemplo, la IA se puede considerar una máquina de predicción útil para asignar de manera eficaz y eficiente los recursos (Agrawal, Gans y Goldfarb, 2018), lo que podría ayudar en la toma de decisiones a los funcionarios encargados de dictar políticas. El objetivo de este manual es centrarse en los temas que son de particular relevancia para el sector público. Las recomendaciones para los funcionarios encargados de dictar políticas también se pueden encontrar en la Recomendación de la OCDE sobre Inteligencia Artificial.[19] A través de este manual se pretende ayudarlos a evaluar las formas en que pueden responder a estas recomendaciones en lo que respecta a la innovación y la transformación del sector público (Recuadro 1.10).

Recuadro 1.10: Respuestas de las políticas a la Recomendación de la OCDE sobre Inteligencia Artificial

La Recomendación de la OCDE sobre Inteligencia Artificial formula cinco recomendaciones a los encargados de dictar políticas nacionales y en materia de cooperación internacional para una IA fiable. Aunque estas recomendaciones van más allá de la innovación y la transformación del sector público, este manual tiene como objetivo ayudar a los encargados de dictar políticas a evaluar posibles aplicaciones en el sector público. Estas recomendaciones y los objetivos correspondientes de este estudio son los siguientes:

- **Invertir en estudios y desarrollo de la IA.** Este manual ofrece información y ejemplos del mundo real que muestran cómo los gobiernos están invirtiendo para

[19] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

Recuadro 1.10: Respuestas de las políticas a la Recomendación de la OCDE sobre Inteligencia Artificial *(Cont.)*

apoyar el uso de la IA en el sector público con miras a transformar e innovar la forma en que diseñan e implementan las políticas y los servicios. Algunos ejemplos son: invertir en la gestión de datos que alimentarán a la IA, así como la apertura de datos para su reutilización a fin de catalizar la innovación.

- **Promover un ecosistema digital para la IA.** En este manual se explica cómo los gobiernos están construyendo puntos de conexión, mecanismos y ecosistemas de intercambio de conocimientos tanto al interior del sector público como con los socios de la industria y la sociedad civil, con el fin de apoyar la experimentación del gobierno en materia de IA.
- **Crear un entorno político propicio para la IA.** Este manual y su respectiva página web "AI Strategies & Public Sector Components" (*Estrategias de IA y componentes del sector público*) (https://oe.cd/aistrategies) documentan la forma en que los países están desarrollando estrategias para la IA y la medida en que prevén específicamente la transformación del sector público. Asimismo, propone un posible marco de referencia para evaluar determinado entorno político a fin de ajustar las actividades para que cumplan con los objetivos de la IA. El OPSI planea integrar en el futuro este recurso con el[20] próximo depósito de conocimientos del Observatorio de Políticas en materia de IA de la OCDE sobre estrategias nacionales de IA, para que los usuarios puedan obtener periódicamente información completa y actualizada sobre éstas.
- **Desarrollar la capacidad humana y preparar la transformación del mercado laboral.** Este manual analiza cómo los gobiernos están desarrollando la capacidad de los funcionarios públicos en torno a la IA. Asimismo, estudia las formas en que los gobiernos están garantizando la capacidad de los socios y proveedores externos, así como la forma en que se están preparando para el potencial de cambios fundamentales en el futuro impulsados por la IA, incluyendo los cambios en el mercado laboral. Además, la respectiva página web "AI Tools & Resources" (*Herramientas y recursos de IA*) (https://oe.cd/airesources) ofrece un depósito categorizado de herramientas y recursos prácticos (por ejemplo, cursos en línea) que pueden ayudar a los funcionarios públicos a aprender más sobre la IA y sus posibles funciones en el sector público.
- **Establecer la cooperación internacional para una IA fiable**. Este manual proporciona un panorama general de las oportunidades de colaboración internacional en desarrollo, incluyendo las convocadas por la OCDE, así como información sobre los productos de dichas colaboraciones, como los principios acordados para una IA fiable.

Fuente: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449; https://oe.cd/aistrategies.

Hoy en día, cumplir con todas las tareas de gobernar puede representar un gran desafío para los gobiernos y las organizaciones públicas que operan en un entorno que evoluciona

[20] http://oecd.ai.

rápidamente, que se enfrentan a problemas cada vez más complejos y a mayores expectativas por parte de los ciudadanos.

Dentro del sector público, la IA podría tener un impacto positivo de diversas maneras. Por ejemplo, podría utilizarse para:

- ayudar a diseñar mejores políticas y tomar mejores decisiones
- mejorar la comunicación y el compromiso con los ciudadanos y residentes
- mejorar la velocidad y la calidad con la que los bienes y servicios públicos se suministran a los ciudadanos
- mejorar el funcionamiento interno de los gobiernos y las organizaciones públicas en general, y ayudar a que los esfuerzos de los funcionarios públicos pasen de las tareas mundanas a un trabajo de alto valor (los estudios demuestran que la IA tiene el potencial de liberar casi un tercio del tiempo de los funcionarios públicos en sólo unos pocos años) (Viechnicki y Eggers, 2017).

Si bien la IA tiene un enorme potencial para generar un efecto positivo en el sector público, lograr estos beneficios no es una tarea fácil. El uso de la IA por parte del gobierno generalmente sigue el mismo camino que el del sector privado (MMC Ventures, 2019), el área y su tecnología asociada es complejo y tiene una curva de aprendizaje pronunciada, y el propósito y contexto dentro del gobierno son únicos y se enfrentan a una serie de retos y otras implicaciones. De hecho, en estudios recientes se ha señalado que "las instituciones gubernamentales, las empresas educativas y las organizaciones de beneficencia están rezagadas en cuanto a la adopción de la IA". Es probable que los miembros vulnerables de la sociedad sean algunos de los últimos en beneficiarse de la IA".[21]

Sin embargo, el OPSI aboga por que el sector público desempeñe un papel importante en el aprovechamiento del potencial de la IA. En algunos casos, el sector ya está desempeñando una función de liderazgo, por ejemplo, mediante la elaboración de principios éticos y procesos de aplicación de la IA basados en los efectos. Para consolidar su posición, es necesario que los funcionarios públicos comprendan el ámbito de la IA y cómo puede repercutir en las políticas y servicios públicos, los empleados del sector público y la forma en que los gobiernos interactúan con su población. También deben aprender de los enfoques, las lecciones aprendidas y los éxitos de los demás. Uno de los objetivos de este manual es ayudarlos a hacerlo. En el Capítulo 2 se analizan los fundamentos técnicos de la IA que deben conocer las partes interesadas del sector público (incluso si la implementación se subcontrata). En el Capítulo 3 se analiza la forma en que los gobiernos están desarrollando estrategias y proyectos de IA. En el capítulo 4 se describe a detalle los factores que los gobiernos deben tener en cuenta al tratar de aplicar la IA en el sector público y se destacan las medidas que pueden adoptar para ayudar a lograr resultados positivos.

[21] www.stateofai2019.com/summary.

¿Qué le depara a la IA?

Se están escribiendo casi tantas tonterías sobre un supuesto e inminente invierno de la IA como sobre una supuesta e inminente explosión de la IA [IA general].

-Yann LeCun, ganador del premio Turing de la ACM, 2018[22]

La IA se está desarrollando rápidamente: la velocidad del cambio ya está afectando a gran parte de la sociedad, incluyendo las interacciones sociales, el trabajo, los sistemas educativos, la organización de la prensa y la forma en que las personas consumen los medios de comunicación, el medio ambiente y así sucesivamente.

Algunos de estos cambios conllevan riesgos conocidos; sin embargo, otros pueden surgir con el paso del tiempo a través de la interacción humana con los sistemas de IA. Existen muchas preguntas sin respuesta e incógnitas con respecto al futuro de la IA. Sin embargo, también hay muchas razones para permanecer optimistas. Varios sistemas de IA del mundo real han demostrado tener resultados significativos, recibiendo a su vez una importante exposición positiva en los medios de comunicación internacionales. En el campo de la medicina, algunos algoritmos de IA han logrado una mayor precisión para identificar tumores que oncólogos experimentados, lo que podría ayudar a reducir las tasas de mortalidad (Nelson, 2019). También se dice que los automóviles autónomos que utilizan IA son más seguros y menos contaminantes que los automóviles conducidos por humanos, dada la información diversa que pueden procesar y su resistencia a la distracción (Meyer, 2019).

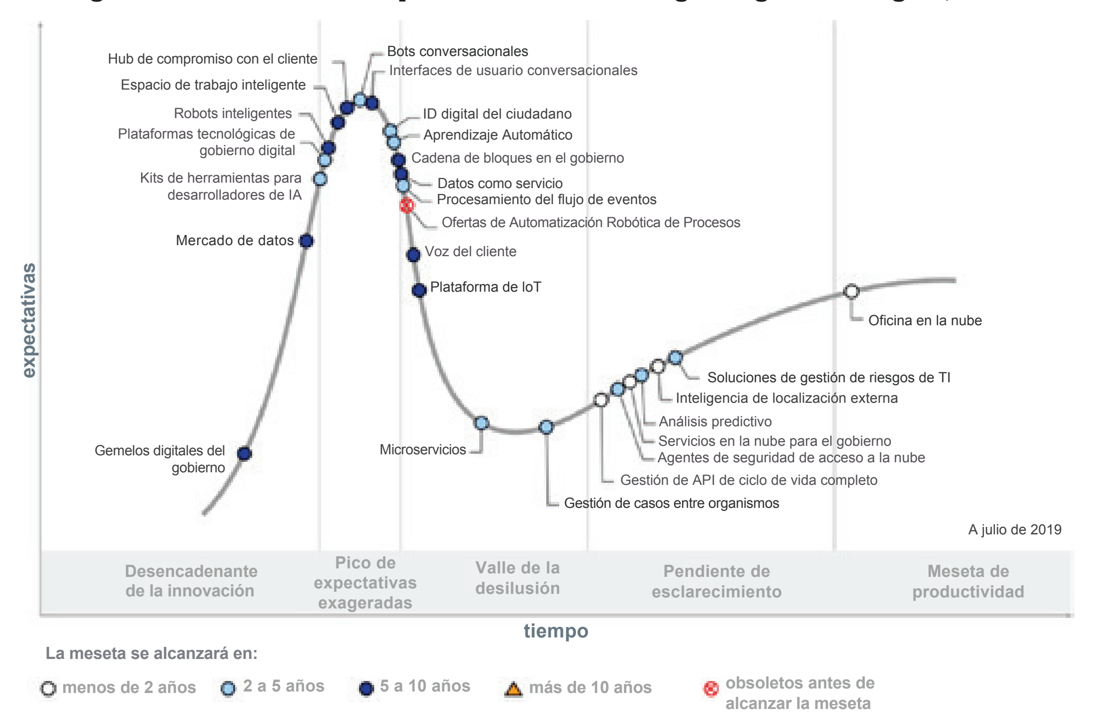
Aun así, como sugiere la historia, se recomienda precaución. El revuelo en torno a la IA puede generar expectativas poco realistas, barreras innecesarias y que se le vea como una panacea en lugar de como una herramienta que puede influir de manera positiva.[23] Varias voces (Marcus, 2018) dentro del mundo de la IA ya están haciendo un llamado a la moderación y a la concientización sobre las actuales limitaciones y desafíos del aprendizaje profundo, una de las áreas prometedoras del desarrollo moderno de la IA (véase el Capítulo 2 para saber más detalles). Las empresas que tratan de sacar provecho de la tendencia de la IA y engañan a los clientes sobre las capacidades o el rigor de la tecnología que se está desarrollando, representan otra amenaza y crean el riesgo de generar expectativas exageradas con resultados reales decepcionantes. Un informe reciente de la empresa de capitales de riesgo londinense MMC, indica que el 40% de las nuevas empresas europeas identificadas como empresas de IA no utilizan realmente la IA de manera significativa para sus negocios (Vincent, 2019).

En lo que respecta al futuro de la IA en el sector público, la situación es un poco menos clara. En comparación con la economía en general y el sector privado, se presta menos atención y se investiga menos sobre la forma en que los ámbitos de investigación y las tecnologías emergentes pueden repercutir y ser utilizados por los gobiernos. Sin embargo, algunas organizaciones han presentado teorías sobre hacia dónde es probable que se dirija la IA en el sector público. Cabe destacar que la empresa mundial de investigación Gartner, predice en su último Ciclo de sobreexpectación de la tecnología de gobierno digital (*Hype Cycle for Digital Government Technology*) (Figura 1.5) que el Aprendizaje Automático, la forma dominante de la IA moderna, ha pasado el "pico de expectativas infladas" y se está deslizando por el "valle

[22] https://twitter.com/ylecun/status/1007099197336760320.
[23] https://correlaid.org/en/blog/problems-ai.

de la desilusión". Sin embargo, también predice que puede tomar sólo de dos a cinco años para que se alcance un estado plenamente productivo, lo cual no es mucho si se consideran las posibilidades de transformación del Aprendizaje Automático.

Figura 1.5: Ciclo de sobreexpectación de la tecnología de gobierno digital, 2019



Fuente: www.gartner.com/en/newsroom/press-releases/2019-08-28-gartner-2019-hype-cycle-shows-cloud-office-has-hit-ma.

Sólo el tiempo dirá si las proyecciones de Gartner son verdaderas. Sin embargo, si los últimos 70 años de historia constituyen un indicador, la IA continuará avanzando y creciendo en todos los sectores, aunque haya periodos de desilusión en el camino. Como resultado, es probable que para los gobiernos no importe si la tendencia es hacia un invierno o una primavera de la IA. Una función importante de los gobiernos democráticos en relación con cualquier tema en evolución es representar legítimamente la voz de su pueblo y encaminar el desarrollo hacia una mejor sociedad en su conjunto. En otras palabras, la IA tiene un gran potencial (tanto positivo como negativo), y es trabajo de los gobiernos garantizar que todas las personas estén en condiciones de cosechar todos los beneficios, y mitigar los riesgos y las consecuencias negativas en su nombre.

Para cumplir con sus funciones, independientemente del momento en que los enfoques de la IA como el Aprendizaje Automático alcancen un estado plenamente productivo, es necesario que los gobiernos y los funcionarios públicos comprendan la IA y cómo puede ésta afectar e influir en el sector público. Lo anterior adquiere cada vez mayor importancia, ya que los gobiernos y sus funcionarios tendrán que tomar decisiones importantes sobre cómo introducir y utilizar la IA en el sector público. De acuerdo con lo que se explica en el Capítulo 3, algunos gobiernos ya están obteniendo beneficios de la IA y prevén un gran potencial para el futuro. En un momento en el que muchos gobiernos se apresuran para ir a

la vanguardia, no hay lugar para los espectadores ni actitudes cautelosas o expectantes. De hecho, no mantenerse actualizado en tecnología conlleva el riesgo de mermar la capacidad del gobierno para encarar problemas cada vez más complejos (IBM Center for the Business of Government, 2019). Es necesario que los gobiernos adquieran un conocimiento experimental en la innovación si quieren seguir tomando decisiones de manera eficaz. Las organizaciones públicas deben comprometerse con la IA y sus tecnologías y enfoques subyacentes, además de considerar las implicaciones para sus instituciones y las interacciones entre la IA y los ciudadanos. Lo anterior es clave para que los funcionarios públicos se familiaricen con las capacidades reales de la IA y para que sus organizaciones se preparen para aprovechar las tecnologías bajo el sistema de IA. Por lo tanto, el OPSI propugna que se experimente con las nuevas tecnologías de manera informada y que se adopten medidas adecuadas para gestionar los riesgos.

El siguiente capítulo de este manual está diseñado para ayudar a los funcionarios públicos (y a cualquier otra persona interesada), a comprender los diferentes enfoques y conceptos técnicos que hay detrás de la IA. En los capítulos siguientes se analiza el contexto del uso de la IA en el sector público y se exponen las consideraciones e implicaciones que pudieran ser más relevantes para el gobierno.

Capítulo 2

Definición de los diferentes métodos de la IA

Cuando recaudas fondos, está la IA. Cuando contratas, está el ML. Cuando implementas, está la regresión logística.

– todos alguna vez en Twitter[1]

Aprendizaje Automático, Redes Neuronales, Aprendizaje Profundo, entre otros términos que se encuentran dentro del ámbito de la Inteligencia Artificial, son cada vez más comunes. Aunque se le atribuye una gran importancia a la IA, muchas de estas conversaciones son en esencia simbólicas debido a la falta de comprensión de la IA. Es común que se perciba a la IA como una "caja negra": aliméntala con datos y ocurrirá algo innovador o futurista. Si bien es cierto que la IA es un campo revolucionario que influye en la sociedad de manera general e impulsa la innovación en todos los sectores e industrias, en muchas ocasiones no se comprende cómo funciona y qué significan realmente los términos relacionados. En el caso de los encargados de dictar políticas y otros funcionarios públicos, esto representa desafíos en cuanto a la forma de regular, apoyar, utilizar y maximizar el posible valor de la IA, al tiempo que se minimizan los aspectos externos asociados que tienen un efecto negativo. En el caso de las empresas, el momento y la forma de utilizar (o no) las tecnologías basadas en la IA son objeto de debate constante. Es fundamental que los responsables de la toma de decisiones en todos los sectores comprendan mejor los beneficios y las limitaciones de la IA, y que sean capaces de reconocer cuando es posible que existan soluciones mejores y más sencillas que hayan sido comprobadas o planteen una mejor solución para el problema que se enfrenta.

Tal y como se describió en el Capítulo 1, la IA significa muchas cosas distintas para muchas personas. Esto se debe en parte a que la IA es un término general para diferentes tipos de enfoques y métodos técnicos. Es importante reconocer que las formas en que se percibe la IA y las diferentes culturas, contextos y antecedentes de quienes la implementan o

[1] https://towardsdatascience.com/no-machine-learning-is-not-just-glorified-statistics-26d3952234e3.

estudian pueden influir en la forma de implementarla, aplicarla o entenderla. Con base en lo anterior, en este capítulo se pretende esclarecer algunos de los diferentes tipos, enfoques y aplicaciones comprendidos en el ámbito de la IA y examinar cómo pueden ser específicamente relevantes para obtener beneficios para el sector público. Si bien no se pretende explicar todas las soluciones o algoritmos técnicos de la IA, en ocasiones la explicación se plantea de forma técnica. Lo anterior es necesario, ya que conocer algunos de los aspectos específicos de la IA puede ayudar a los líderes del gobierno y a los funcionarios públicos de muchas maneras. Por ejemplo, comprender algunos de los fundamentos técnicos puede ayudar a los funcionarios a decidir cuál de esas herramientas es más útil para resolver problemas concretos, y a estar más preparado para las negociaciones con los vendedores que tratan de vender la IA o soluciones más tradicionales.

Sin embargo, antes de empezar es importante reconocer una condición previa importante de casi todos los proyectos de IA: datos de calidad. Los expertos suelen plantear la importancia de los datos de calidad (o la falta de éstos) como un factor importante que contribuye al éxito o fracaso de una iniciativa de IA. Algunos incluso afirman que la mayoría de los gobiernos simplemente no están preparados para utilizar la IA y deberían enfocarse primero en poner en orden sus datos (Schneider, 2019). ¿Pero qué es lo que implica?

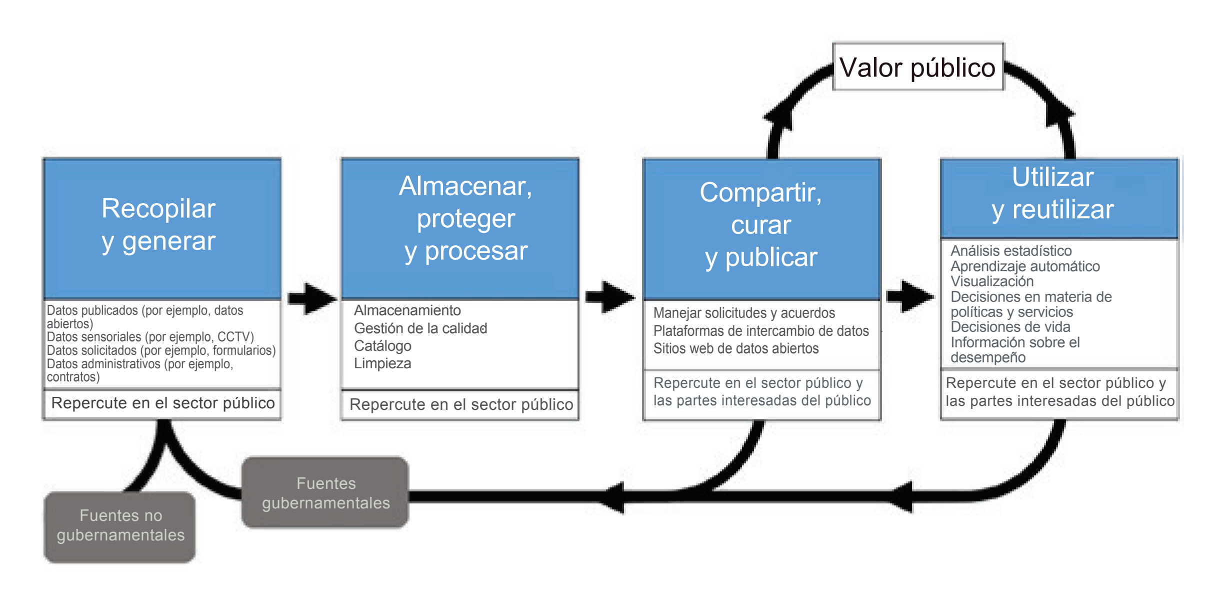
Datos: el alimento de la IA

La mayoría de los proyectos de IA involucran datos como una entrada crítica. Lo anterior ocurre en particular en los proyectos de Aprendizaje Automático, en los cuales el objetivo es aprender de los datos. Sin embargo, no todos los datos son iguales y se deben tomar medidas para garantizar que los datos utilizados para un proyecto de IA sean precisos, confiables y adecuados para la tarea en cuestión. Incluso cuando la IA podría constituir una solución a los problemas gubernamentales, es posible que la falta de técnicas básicas de gestión de datos limite su potencial. Los funcionarios públicos que están interesados en interactuar con la IA necesitan saber qué son los datos, qué tipos de datos se pueden utilizar, qué tipo de datos necesita la IA y cómo verificar si sus datos están listos para ésta.

Para comprender mejor de qué forma los datos pueden y deben sentar las bases para implementar la IA, el Ciclo de Valor de los Datos Gubernamentales (Figura 2.1) ilustra el ciclo de vida de los datos y cómo los gobiernos pueden utilizarlo para generar valor público. El ciclo incluye proporcionar valor a través de técnicas de IA, tales como el Aprendizaje Automático, pero también a través de otros medios no relacionados con ella. En el documento de trabajo de la OCDE *A Data-Driven Public sector: Enabling the Strategic Use of Data for Productive,* Inclusive and Trustworthy Governance (*Un sector público basado en datos: Habilitar el uso estratégico de datos para una gobernanza productiva, inclusiva y confiable*) (van Ooijen, Ubaldi y Welby, 2019) se abordan con más detalle estos temas. Además, el informe del OPSI sobre *Fostering Innovation in the Public Sector* (Cómo impulsar la innovación en el sector público) (OCDE, 2017a) incluye un capítulo dedicado a la "Gestión de datos, información y conocimientos para la innovación". Se invita a los interesados en estudiar los fundamentos de la IA relacionados con los datos, a que lean las publicaciones antes mencionadas, ya que este manual ofrece un análisis más superficial sobre esta área.[2]

[2] Además, la Hoja de puntaje de preparación organizativa en el marco de madurez de los datos (*Data Maturity Framework Organisational Readiness Scorecard*) de la Universidad de Chicago, es útil para evaluar mejor su desempeño actual en la gestión de datos e identificar las áreas en las que deben mejorar. Véase https://dsapp.uchicago.edu/home/resources/datamaturity.

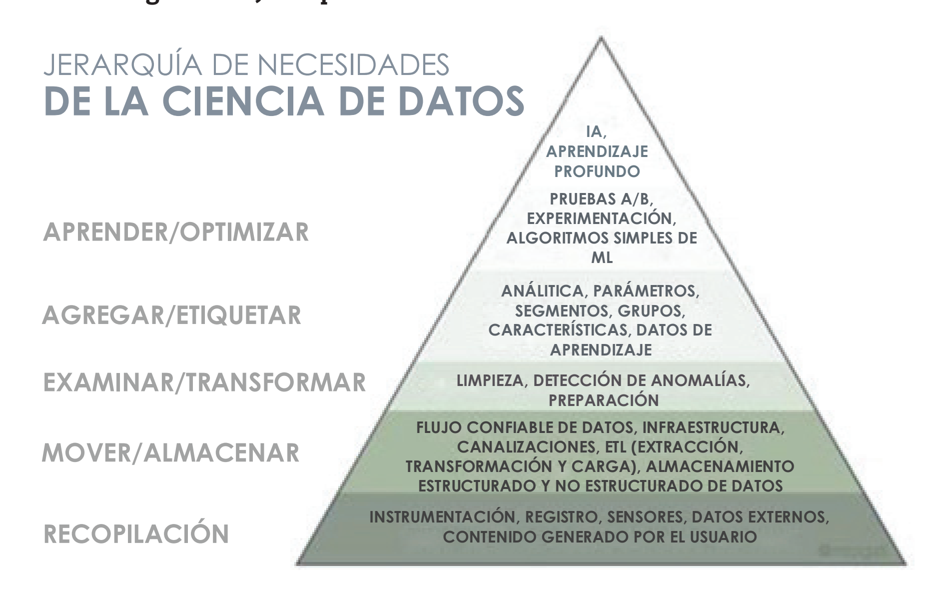
Figura 2.1: Ciclo de valor de los datos gubernamentales



Fuente: van Ooijen, Ubaldi and Welby (2019), *A Data-Driven Public Sector: Enabling the Strategic Use of Data for Productive, Inclusive and Trustworthy Governance*, OECD Working Papers on Public Governance, No. 33, OECD Publishing, París, https://doi.org/10.1787/09ab162c-en.

Otra visualización reconocida entre los científicos de datos y los especialistas de la IA es la "Jerarquía de necesidades en la ciencia de datos", elaborada por la científica de datos Monica Rogati. En ella se busca explicar por qué las organizaciones necesitan una base de datos sólida para lograr una IA y un Aprendizaje Automático eficaces.

Figura 2.2: Jerarquía de necesidades de la ciencia de datos



Fuente https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007.

La Jerarquía de necesidades de la ciencia de datos comprende cinco niveles principales y se lee de abajo hacia arriba según su importancia. Aunque se presenta a manera de jerarquía, el desarrollo de un proyecto de IA no es un proceso meramente lineal. Por el contrario, la pirámide debe servir para enfatizar la importancia de tomar decisiones fundamentales en las primeras etapas del proceso.

Tanto en el caso del ciclo de valor de los datos gubernamentales como en la jerarquía de necesidades de la ciencia de datos, será necesario volver a revisar los pasos anteriores y adaptarse a las nuevas circunstancias a medida que se disponga de nueva información o estrategias. Por consiguiente, no deben considerarse como un proceso rígido y fijo sino como un proceso dinámico e iterativo. Es fundamental que los gobiernos consideren algunas de las áreas de interés que tienen en común a fin de crear una gestión e infraestructura de datos sólidas. En el análisis que se presenta a continuación se plantean los aspectos que en términos generales se requieren para sentar las bases que posibiliten la experimentación y la aplicación de la IA. Para cimentarlas, será esencial que los gobiernos inviertan en capacidades y recursos y así ascender, en sentido metafórico, en la jerarquía de necesidades.

Además de esta sección, en el Capítulo 4 se analizan las consideraciones y las prácticas gubernamentales específicas necesarias para recopilar, acceder y utilizar de manera ética los datos de calidad. El Anexo A incluye un estudio de caso sobre la estrategia y hoja de ruta del gobierno de los Estados Unidos en materia de datos (*United States Federal Data Strategy and Roadmap*), en el cual se ilustra cómo un país está tratando de crear una base de datos sólida para la IA.

Recopilación: ¿Qué datos se necesitan? ¿existen?

Al considerar el uso de un sistema de IA para resolver un problema específico, es importante determinar qué tipos de datos existen (tanto en el sector público como en el privado) y si están disponibles para utilizarse. ¿Son suficientes los datos disponibles para resolver el problema o se necesitarán datos adicionales? Además, ¿se puede acceder a los datos? Algo que debe considerarse a este respecto es que los datos siempre son el resultado de un proceso de recopilación específico con un propósito definido. Tal y como se mencionó anteriormente, tanto la IA como su aplicación pueden verse afectadas por contextos, tradiciones y creencias. Los procesos de recopilación de datos también están sujetos a dichas interpretaciones e influencias y requieren una atención especial, ya que el proceso de recopilación nunca es imparcial. Se puede decir que los datos no existen hasta que se recopilan, y las decisiones sobre *cómo* recopilarlos influyen en última instancia en los datos que *se* recopilan y generan. Asimismo, es fundamental analizar *qué* tipo de datos se recopilan, ya que el hecho de decidir sobre qué datos se recopilan implica hacer una declaración implícita de valor sobre lo que es importante, y tiene consecuencias en aquello que puede obtenerse a partir de los datos.

Existen muchas formas diferentes de recopilar o generar datos. Los datos cualitativos pueden obtenerse mediante métodos como entrevistas individuales, grupos de discusión, observaciones de campo o investigaciones activas, mientras que los datos cuantitativos pueden recopilarse mediante diversos experimentos o mediante la realización de encuestas y la administración de cuestionarios. Los gobiernos suelen tener grandes cantidades de datos a su disposición, los cuales se han recopilado para diversos fines, o pueden obtener dichos datos de otros gobiernos mediante programas de Datos Abiertos Gubernamentales (DAG). Los gobiernos también pueden adquirir u obtener datos de otra manera (por ejemplo,

mediante acuerdos o incentivos de intercambio de datos) del sector privado y la sociedad civil. Independientemente de la fuente y el tipo de datos, es importante procurar que los datos sean legibles por máquina y vayan acompañados de metadatos de alta calidad.[3] En el Capítulo 4 se analizan otras consideraciones y orientaciones para garantizar la recopilación y el acceso éticos a los datos.

En algunas partes de esta sección se utiliza un conjunto de datos a manera de ejemplo. La tabla 2.1 es un extracto del conjunto de datos de la *Flor iris* introducido por el biólogo Ronald Fisher en 1936. El conjunto de datos flor iris es un ejemplo que se utiliza comúnmente en el campo de la IA y en especial en el Aprendizaje Automático.

Tabla 2.1: Un extracto del *conjunto de datos flor iris*

Registro	Longitud del pétalo	Ancho del pétalo	Longitud del sépalo	Ancho del sépalo	Especie
1	5.1	3.5	1.4	0.2	*Iris setosa*
2	4.9	3.0	1.4	0.2	*Iris setosa*
…					
150	5.9	3.0	5.1	1.8	*Iris virginica*

El conjunto de datos consiste en un registro de varias flores de iris observadas en la naturaleza con sus dimensiones morfológicas (longitud y ancho del pétalo y del sépalo) y la especie particular de iris a la que pertenecen. En el Aprendizaje Automático, es un ejemplo de un esfuerzo por predecir la especie de una flor con base en sus dimensiones.

Fuente: Fisher, R.A. (1936), "The use of multiple measurements in taxonomic problems", *Annals of Eugenics*, Vol. 7/2, pp. 179-188.

Almacenamiento, protección y habilitación del flujo de datos

Para poder aprovechar diferentes fuentes de datos y así alimentar la IA es necesario reconsiderar la forma en que se almacenan los datos, la manera en la que fluyen y cómo están estructuradas las organizaciones, de qué forma se gestiona el trabajo y cómo las personas están conectadas y en red. Cada vez se ejerce más presión para que se invierta en una infraestructura moderna de TI, y en mejoras a la interoperabilidad y los metadatos. Los gobiernos deben contar con estrategias sólidas de gestión de datos para lograrlo y garantizar que los datos se puedan reconocer y estén disponibles a partir de esta amplia gama de fuentes dentro y fuera del sector público, de manera accesible y oportuna (OCDE, 2015b).

Las burocracias jerárquicas tradicionales suelen tener flujos horizontales limitados de datos debido a la rigidez de las regulaciones y prácticas incompatibles de gestión de la información o rivalidades y competencia entre organizaciones anticuadas (OCDE, 2017a). En algunos casos, las dependencias gubernamentales se enfrentan a dificultades para intercambiar datos, ya sea por obstáculos técnicos, barreras administrativas o ambos. Incluso puede resultar difícil intercambiar datos dentro de las dependencias. Por lo tanto, los avances de la

[3] Para leer un resumen y análisis más detallado sobre la legibilidad por máquina, véase la guía en www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data.

IA en el sector público serán limitados, a menos que los datos puedan fluir fácilmente y, en última instancia, se empleen de manera que alimenten los algoritmos y resuelvan los problemas. Para facilitar este proceso, deben eliminarse las barreras y crear mecanismos que permitan este flujo. En Europa, la Comunidad de Interoperabilidad Semántica (SEMIC, por sus siglas en inglés) está desarrollando un acuerdo común sobre los datos por medio del cual se facilite el intercambio entre las administraciones públicas europeas y se permita el suministro de servicios digitales transfronterizos e intersectoriales.[4] Análogamente, la Declaración de Tallin sobre gobierno electrónico de 2017[5] estableció un compromiso político para la implementación del “Principio de solo una vez” (*Once Only Principle*), de acuerdo con el cual los estados miembros de la UE “se esforzarán para crear una cultura de reutilización, incluyendo la reutilización responsable y transparente de los datos dentro de nuestras administraciones” (EU2017.ee, 2017). Por consiguiente, será necesario crear nuevos mecanismos para permitir la libre circulación de los datos. Un ejemplo de este mecanismo se puede encontrar en Canadá, donde el gobierno ha creado una API Store, una “ventanilla única de API” del sector público.[6] Dichas herramientas permiten el desarrollo de algoritmos y aplicaciones basados en la IA que consumen datos a través de estas API. Sin embargo, aún quedan por resolver varios problemas culturales, técnicos y en materia de procedimientos, tal como se describe en los informes de la publicación *Data-driven Public Sector* and *Fostering Innovation in Government* (*Sector público basado en datos* y *Cómo impulsar la innovación en el gobierno*).

Uno de los aspectos que deben tomar en cuenta los organismos gubernamentales en lo que respecta al almacenamiento y uso de datos para la IA es el mantenimiento y continuidad de éstos. Es importante que los datos que se utilizan para la IA estén disponibles de manera constante y coherente y en el mismo formato o estructura al que está acostumbrado el algoritmo. Dado que es posible que las organizaciones gubernamentales y las prácticas de recopilación de datos cambien, podría darse el caso que los datos de los que dependían anteriormente los algoritmos ya no se recopilen o se proporcionen en un formato diferente al requerido. Para evitar que esto suceda, los organismos gubernamentales que implementan la IA deben tener conocimiento de los diferentes proyectos de IA en marcha, los algoritmos utilizados y los datos de los que dependen. Esto ayudaría a impedir que se cancele un proyecto de IA en el que se invirtió tiempo y esfuerzo debido a cambios en la disponibilidad de datos. Una estrategia que comúnmente utilizan los organismos para evitar dichas situaciones es creando diccionarios de datos o catálogos de datos que incluyan información arquitectónica, información de metadatos, tipo de datos, fuente de datos y otra información relevante. Esta información podría mantenerse actualizada en los portales de datos abiertos gubernamentales.

La seguridad es también una consideración importante. Como se señala en un documento de trabajo de la OCDE (van Ooijen, Ubaldi y Welby, 2019), cuando los datos se almacenan y procesan, es fundamental que las bases de datos y los acuerdos sobre cómo se gestionan sean transparentes. Los datos “inactivos” en esta etapa suelen ser los más vulnerables a las amenazas contra la seguridad digital. Es interesante que al comprometerse con niveles más altos de interoperabilidad y asegurarse de que las copias de los datos no estén en manos de múltiples organismos, el costo de proteger los datos tienda a disminuir, lo que

[4] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/about.
[5] https://ec.europa.eu/newsroom/document.cfm?doc_id=47559.
[6] https://api.canada.ca.

puede ayudar a incrementar las capacidades de seguridad cibernética de los gobiernos. Los gobiernos también pueden considerar en un inicio el uso de "espacios aislados" (*sandbox*) para experimentar y desarrollar la IA, lo cual puede servir para separar los datos mientras los funcionarios aprenden sobre su potencial y respuesta (CIPL, 2009) (en el Capítulo 4 se explican con más detalle los espacios aislados). La seguridad de la información y los datos es un campo completo en sí mismo y está fuera del ámbito de este manual. Sin embargo, es necesario que los organismos del sector público inviertan tiempo y recursos para asegurarse de que comprenden esta área y cuentan con las medidas de seguridad adecuadas.

Transformación: Preparando los datos para su uso

> *No puedes comparar manzanas y naranjas. Bueno, en realidad, sí puedes. Sólo se necesita preparación.*

Una vez que se cuente con los datos, es importante asegurarse de que sean de buena calidad para realizar análisis significativos. Este paso es, al mismo tiempo, uno de los más importantes, más subestimados, ignorados y de los que menos se disfrutan. Es un proceso intensivo y que requiere mucho tiempo. Según una encuesta realizada a científicos de datos reportada por *Forbes*,[7] se estima que los científicos de datos dedican hasta un 60% de su tiempo a tareas relacionadas con la transformación de los datos, un 20% a su recopilación y sólo un 20% al análisis real.

Para abordar la metáfora de "manzanas y naranjas", ambas se pueden comparar. Estos dos tipos de frutas tienen formas similares y colores diferentes. Tienen una textura diferente, un sabor diferente y se pueden cosechar en el mismo país. Se puede establecer un precio para cada tipo. En otras palabras, el análisis es posible cuando se establecen los parámetros para el mismo.

Esta transformación implica la *limpieza de los datos*[8] (también denominada "data wrangling" o "munging" [*preparación de los datos*]), un "proceso de exploración y transformación iterativas de datos que facilita el análisis".[9] Por ejemplo, es posible que se hayan recopilado datos sobre las manzanas y las naranjas, pero el color podría describirse utilizando un texto en un caso (por ejemplo, rojo, verde) y un valor numérico en el otro (por ejemplo, 1 = rojo, 2 = verde). Es posible que haya información sobre los precios de las manzanas, pero no de las naranjas. Por error, alguien pudo haber incluido datos sobre los plátanos, que están fuera del ámbito de este análisis específico. El Recuadro 2.1 que se muestra a continuación, proporciona un resumen de los problemas que se pueden encontrar comúnmente durante la limpieza de datos y sus nombres técnicos.

7 www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#5dd8ff3a6f63.
8 https://hci.stanford.edu/courses/cs448g/lectures/CS448G-20110411-DataCleaning.pdf.
9 https://web.stanford.edu/class/cs442/lectures_unrestricted/cs442-visualization.pdf.

Recuadro 2.1: Problemas que comúnmente se observan al limpiar los datos

El ejemplo del conjunto de datos de la flor iris se utiliza para ilustrar los problemas comunes que implica la limpieza de datos.

Registro	Longitud del pétalo	Ancho del pétalo	Longitud del sépalo	Ancho del sépalo	Especie
1	N/A	3.5	1.4	0.2	*Iris setosa*
2	4.9	3.0	Long	0.2	*Rosa dumalis*
…					
150	5.9	200.3	5.1	1.8	*Iris virginica*

Valores faltantes (en rojo): Algunas veces es posible que determinados valores de los conjuntos de datos no estén disponibles (N/A, no disponible o NULO). El hecho de que un conjunto de datos esté incompleto *puede* deberse a un problema en el proceso de recopilación de datos, a una mala manipulación posterior o al resultado de un acto malintencionado. Si se trata de un problema de recopilación de datos, los valores faltantes pueden ser perjudiciales para los algoritmos y generar predicciones erróneas, ninguna predicción o incluso dañar el sistema de IA. Para evitar los resultados negativos de los valores faltantes, se pueden adoptar medidas para comprender la causa de la falta de datos. Entre las soluciones fáciles se pueden mencionar la sustitución del valor ausente por un valor predeterminado o la extrapolación a partir de los valores existentes. Existen técnicas más sofisticadas; sin embargo, no se incluyen en este documento.

Valores atípicos (en naranja): Un valor atípico de un conjunto de datos es un punto de datos que se desvía de los demás. Las causas de los valores atípicos pueden ser las mismas que las de los valores faltantes. También pueden indicar un inconveniente en el modelado del problema, como el hecho de no haber considerado un parámetro o fenómeno importante. Sin embargo, un valor atípico también puede indicar un punto de datos u objeto de estudio interesante. Al igual que en el caso de los valores faltantes, los valores atípicos derivados de errores o problemas pueden causar dificultades y afectar el análisis de forma negativa. En el ejemplo anterior, el ancho del pétalo de una flor tiene más de 200 cm, mientras que el de todas las demás iris es de entre 0 y 5 cm. Cuando se calcula el ancho promedio del pétalo, la inclusión o exclusión de un valor atípico tiene un efecto importante y puede ser objeto de muchas interpretaciones durante el análisis posterior.

Valores inesperados (en azul): En el conjunto de datos de la flor iris, el valor *largo* para la longitud del sépalo se encuentra en un formato diferente al esperado (un texto en lugar de un número), mientras que el valor *Rosa dumalis* para las especies está fuera del intervalo de opciones disponibles (una rosa en lugar de iris). También en este caso, muchas causas podrían explicar la aparición de valores inesperados. Una vez que se haya examinado y determinado que son resultado de un error o problema, es necesario corregir estos valores eliminando el registro, investigando el proceso de recopilación o sustituyendo el valor.

Fuera del contexto de la limpieza de datos, analizar las causas de este tipo de cuestiones puede resultar beneficioso. Por ejemplo, la detección de valores atípicos y anomalías puede tener aplicaciones en sectores como el bancario y financiero (véase *Aprendizaje no supervisado* y *Agregación* más adelante en este capítulo).

Fuente: OPSI.

También puede diferir el formato en que se han recopilado los datos: es posible que los datos de las manzanas se hayan ingresado correctamente en una tabla en un archivo Excel o CSV *(datos estructurados)*, y que los datos de las naranjas incluyan, por el contrario, imágenes y notas manuscritas *(datos no estructurados)*. En la práctica, existen formatos de archivo para datos como CSV, XML y JSON, y pueden utilizarse diferentes sistemas para gestionar las bases de datos (por ejemplo, SQL para *bases de datos estructuradas* y NoSQL para *bases de datos no estructuradas*).

En esta etapa (y la siguiente), es importante considerar si los datos son vulnerables al sesgo y, de ser así, tomar medidas para mitigarlo. Más adelante en este capítulo y en el Capítulo 4 se analiza con más detalle el sesgo de datos.

Los pasos para determinar dichas discrepancias en los datos y tomar decisiones sobre cómo resolverlas son subestimados y pueden llegar a considerarse tareas fáciles o secundarias. De hecho, estos pasos son esenciales para generar análisis sólidos. Una falta de comprensión de los datos puede llevar a resultados iniciales incorrectos y, por lo tanto, a tomar decisiones poco sensatas. Una vez que se hayan hecho las observaciones y tomado decisiones respecto a los datos disponibles, se puede llevar a cabo una limpieza de los datos per se para resolver los problemas. Ello implica realizar las modificaciones pertinentes a los datos, un proceso que puede acelerarse mediante el uso de tecnologías de IA basadas en reglas, como la Automatización Robótica de Procesos (la cual se explica con mayor detenimiento más adelante en este capítulo).

Agregar y comprender los datos

El penúltimo paso antes de comenzar a experimentar con la IA en un proyecto de datos es agregar y analizar los datos para comprenderlos y prepararlos mejor para la IA. Esta etapa es vital, ya que interpretar y comprender los datos de los que se dispone representa una proporción significativa del tiempo y los recursos necesarios para diseñar sistemas de IA, incluso más que la creación de los modelos y algoritmos. En esta etapa del proceso, es probable que ya se haya llegado a algunas conclusiones preliminares sobre los datos, pero aún es necesario trabajar para generar hipótesis de trabajo para las pruebas.

Por ejemplo, es posible que se haya recopilado una cantidad importante de información a partir de las entrevistas, pero es necesario estructurar las notas realizadas. Otra posibilidad es que una empresa de telefonía celular haya concedido acceso a sus datos sobre itinerancia y haya ayudado a establecer un protocolo de intercambio de datos; sin embargo, quizá ahora sea necesario combinarlo con datos geográficos para ayudar a mejorar la gestión del flujo de tráfico o los servicios de transporte público. En ambos casos, se necesita realizar un análisis minucioso del problema en cuestión y de cómo podrían ayudar las variables específicas de los datos.

Los pasos clave del proceso son el *etiquetado, las anotaciones* y *la ingeniería de características* (en el Recuadro 2.2 se proporcionan más detalles sobre algunos de estos términos clave). Es importante prestar atención a los metadatos asociados para garantizar que los datos que se están analizando se hayan comprendido correctamente. Tanto el etiquetado como las anotaciones implican observar los datos y colocar una etiqueta o anotación adicional para garantizar una mayor facilidad de uso, claridad y legibilidad. En el caso de los datos visuales, como las imágenes, puede ser necesario que se clasifique o etiquete el contenido de la

imagen (por ejemplo, si se trata de la imagen de una naranja o una manzana). El etiquetado, en particular, implica identificar los elementos que un sistema de IA tratará de predecir. La ingeniería de características es el proceso de convertir determinada cantidad de datos en una característica u otro tipo de datos que pueden ser consumidos por un proyecto de IA.

Una forma de abordar este análisis es definiendo los criterios del éxito para el problema en cuestión, y posteriormente, partiendo del éxito, determinar los parámetros o los indicadores clave de desempeño (KPI, por sus siglas en inglés) que deben estar en un cuadro de mando a fin de determinar si se ha alcanzado el punto de éxito. Una acción clave es identificar una medida única de éxito para que se pueda rastrear el proyecto a lo largo del tiempo *(etiqueta)* y otros aspectos que podrían influir en él *(características)*. En algunos casos, dichos elementos no son evidentes de inmediato. El uso del *Aprendizaje Automático no supervisado* (una forma de IA que se analizará más adelante en este capítulo) puede ayudar a evaluar los datos y su estructura a fin de seleccionar mejor las características, tal como se describe en la sección que trata este tema más adelante.

Recuadro 2.2: Términos clave para analizar y etiquetar los datos para la IA

Registro	Longitud del pétalo	Ancho del pétalo	Longitud del sépalo	Ancho del sépalo	Especie
1	5.1	3.5	1.4	0.2	*Iris setosa*
2	4.9	3.0	1.4	0.2	*Iris setosa*
…					
150	5.9	3.0	5.1	1.8	*Iris virginica*

La **etiqueta** es la variable dependiente. Esta es la variable cuyo valor está tratando de predecir el proceso. También se puede denominar *respuesta*, *resultado* o *salida*. En este caso, la *especie* es la etiqueta y el objetivo es predecir el valor *(setosa*, *virginica* o *versicolor*). La *etiqueta* es de particular importancia en el contexto del *aprendizaje supervisado,* un tipo de Aprendizaje Automático (véase la sección que trata este tema más adelante).

Una **característica** es una variable independiente en un conjunto de datos. Esta es una de las variables que se utilizan para hacer predicciones. En este caso, el objetivo es predecir la especie de iris con base en sus dimensiones y los factores descriptivos de la tabla de longitud y ancho de los pétalos y sépalos.

La ingeniería de características, según el glosario de Aprendizaje Automático de Google, es "el proceso de determinar qué características serían útiles para que un modelo aprenda". En otras palabras, la ingeniería de características se centra en seleccionar los datos correctos que deben guardarse como *características* para hacer predicciones e identificar la etiqueta.

La selección de características es sólo un aspecto de la ingeniería de características. Es posible que en algunas ocasiones los datos existan, pero no en el formato requerido por el algoritmo. Por ejemplo, puede ocurrir que el modelo requiera la edad de una persona para que funcione el cálculo, pero sólo se dispone de la fecha de nacimiento.

Recuadro 2.2: Términos clave para analizar y etiquetar los datos para la IA *(Cont.)*

Aquí, es necesario **normalizar** los datos, que en este caso se entiende como la conversión de los valores de su estado *sin procesar* a un intervalo estándar de valores.

Las computadoras suelen funcionar mejor cuando tratan con números. La *característica de color* de la flor del conjunto de datos de la flor iris se pudo haber añadido utilizando un código de color en lugar de texto (por ejemplo, rojo es 1, azul es 2, amarillo es 3). De igual manera, es posible que la edad de una persona sea menos relevante y, por lo tanto, se indique mediante un umbral ("la persona tiene 30 años o menos" = 0, "la persona tiene más de 30 años" = 1).

Fuente: http://archive.ics.uci.edu/ml/datasets/Iris, https://developers.google.com/machine-learning/glossary; https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7; www.kdnuggets.com/2019/02/quick-guide-feature-engineering.html.

Una vez que se hayan identificado las características y las etiquetas, se pueden comenzar a desarrollar datos de aprendizaje que permitan crear algoritmos de Aprendizaje Automático. En la sección *Aprendizaje automático 101* que se incluye más adelante en este capítulo, se proporciona más información sobre los algoritmos de aprendizaje y análisis.

Implementación

Una vez que los datos se hayan recopilado, comprendido y transformado en un formato adecuado, la etapa final de la Jerarquía de necesidades de la ciencia de datos es aplicar la IA para ayudar a generar algunas respuestas a las preguntas de interés del equipo del proyecto. Tal como se explica a detalle más adelante en las siguientes secciones, los sistemas sencillos de IA se basan en una combinación de múltiples reglas del tipo SI…ENTONCES, mientras que en los métodos modernos se prefiere el Aprendizaje Automático y las técnicas de aprendizaje profundo en lugar de los métodos basados en reglas o en combinación con estos.

Independientemente del enfoque que se adopte o del algoritmo específico, lo importante es experimentar con las diferentes técnicas iniciando con las más simples, comparar sus resultados con respecto que se requería resolver al inicio o las hipótesis que se probarán, y mejorar gradualmente las técnicas o avanzar hacia modelos más complejos si los resultados no son satisfactorios.

Las siguientes secciones exponen algunos conceptos básicos sobre cómo funcionan los diferentes tipos de IA para poder determinar qué tipo de IA es el adecuado para el problema en cuestión. También se presentan ejemplos de la IA que utilizan algunas empresas y organizaciones públicas ya que es posible que diferentes organizaciones se enfrenten a situaciones similares.

Evolución de la IA: IA basada en reglas frente al Aprendizaje Automático

En la sección anterior se explicó que los datos son un requisito previo fundamental para la IA. En esta sección se exploran los diferentes tipos de IA y los diversos métodos que se pueden utilizar para desarrollar sistemas de IA.

A los efectos del presente informe, la IA puede dividirse en dos tipos diferentes: 1) IA basada en reglas (también conocida como "IA simbólica"), y 2) IA que emplea el Aprendizaje Automático. Con el tiempo, el Aprendizaje Automático se ha convertido en el modelo dominante, y la cuestión de si la IA ha evolucionado hasta un punto en el que sólo el método del Aprendizaje Automático debería considerarse "inteligente" es objeto de debate entre los expertos, los profesionales y las empresas de IA. Ambas pueden ser relevantes para el sector público y ofrecidas por los vendedores, por lo que se incluyen en este documento. Además, algunos sistemas híbridos de IA emplean tanto métodos basados en reglas como de Aprendizaje Automático. Por ejemplo, el Aprendizaje Automático se puede utilizar para detectar correlaciones y patrones que posteriormente se utilizan para determinar qué tipos de reglas funcionarían mejor en un sistema basado en reglas (Dickson, 2019; Mao et al., 2018).

Figura 2.3: La relación entre la Inteligencia Artificial y el Aprendizaje Automático

Fuente: OCDE (2019), *Artificial Intelligence in Society*, proporcionada por la Iniciativa para la Investigación de Políticas de Internet (*Internet Policy Research Initiative*, IPRI) del Instituto Tecnológico de Massachusetts (MIT).

En este capítulo se explican brevemente las características de cada tipo de IA, incluyendo la forma en que se diferencian entre sí, sus fortalezas y limitaciones relativas. En el Recuadro 2.3 se destacan las diferencias entre los sistemas basados en reglas y los sistemas de Aprendizaje Automático, utilizando el ejemplo de una computadora que aprende a jugar al ajedrez.

Recuadro 2.3: Cómo enseñar a una computadora a jugar al ajedrez: IA basada en reglas y Aprendizaje Automático

A lo largo de la historia de la IA, el ajedrez se ha utilizado para ilustrar cómo funcionan los diferentes métodos y evaluar sus habilidades. Un tablero de ajedrez consiste en 64 cuadros, dispuestos en una cuadrícula de 8x8 con un patrón de cuadros blancos y negros. Hay dos posibles resultados en una partida de ajedrez: uno de los jugadores gana o se produce un empate.

Recuadro 2.3: Cómo enseñar a una computadora a jugar al ajedrez: IA basada en reglas y Aprendizaje Automático *(Cont.)*

Basada en reglas

Cada pieza de ajedrez tiene sus propias reglas específicas que determinan los movimientos que puede hacer en el tablero. En seudocódigo, esta regla podría programarse como:

- "**SI** la pieza es un peón**, ENTONCES** sólo puede moverse una posición hacia adelante en línea recta"
- "**SI** la pieza es una reina**, ENTONCES** puede moverse en todas las direcciones y a lo largo de los cuadrados que estén libres"

Los jugadores sólo pueden realizar un movimiento durante su turno:

- "**SI** es turno del jugador 1**, ENTONCES** permitir que mueva una pieza"

Los jugadores pueden comerse la pieza de un oponente moviéndose a un espacio ocupado y así sucesivamente.

Todas las reglas del ajedrez se pueden formular utilizando la estructura "SI… ENTONCES" **(SI** ocurre determinada condición **ENTONCES** se produce determinada acción), las cuales, en conjunto, describen el juego completo. Este conjunto de reglas se conoce como la "base de conocimientos" del sistema.

Con base en esto, se pueden desarrollar muchos conjuntos de reglas del tipo SI… ENTONCES para calcular todos los movimientos permitidos en una situación dada, así como el mejor movimiento que se puede realizar, dado un número de situaciones previamente programadas para ganar la partida.

Aprendizaje Automático

El método de Aprendizaje Automático del ajedrez tiene un punto de partida diferente. En lugar de tratar de enumerar de forma explícita todas las reglas del juego, se recaban datos sobre un número significativo de partidas de ajedrez que se jugaron anteriormente. Los datos podrían incluir los movimientos realizados por los jugadores, el resultado de un movimiento (si se comió una pieza o no) y el resultado general del juego (ganar, perder o empatar). Cada juego es diferente debido a las posibles combinaciones de movimientos de los jugadores contrarios.

Una vez que se alimenta al sistema de la IA con estos datos, es capaz de aprender a inferir las reglas del juego por sí mismo (en lugar de estar preprogramado con reglas y movimientos óptimos) y, posteriormente, se pueden utilizar para jugar nuevas partidas.

Nota: Para obtener más información sobre la creación de una IA de ajedrez sencilla consultar: https://medium.freecodecamp.org/simple-chess-ai-step-by-step-1d55a9266977 y www.j-paine.org/students/lectures/lect3/node5.html.

IA basada en reglas

La IA basada en reglas, si es que puede considerarse IA, se denomina también "programación clásica", "razonamiento simbólico" o "IA simbólica", "sistemas expertos" o incluso "inteligencia artificial a la antigua" (*Good Old-Fashioned Artificial Intelligence*) (Haugeland, 1985). Tal y como se describió, este tipo de sistemas se componen de una sucesión de reglas del tipo SI…ENTONCES programadas por humanos para describir una lógica de negocio o flujo de trabajo. En caso de considerarse como IA, son el tipo más simple de IA. Debido a su simplicidad conceptual, los sistemas basados en reglas suelen tener un alto nivel de *interpretabilidad* y *explicabilidad*.[10] De hecho, la lógica detrás de las estructuras del tipo SI…ENTONCES es lo suficientemente explícita como para que, sin conocimiento previo, cualquiera pueda leer las diferentes reglas por separado y comprenda o "interprete" lo que el sistema está haciendo.

Con sólo unas cuantas reglas adecuadas, es posible crear un sistema más o menos elaborado que sea aplicable a muchas situaciones diferentes. Por ejemplo, el caso del ajedrez que se analizó anteriormente demuestra que se puede utilizar un conjunto limitado de reglas para crear un juego complejo que describa muchas situaciones posibles diferentes. En la Figura 2.4 se presenta un diagrama de flujo de decisiones para un árbol de menús telefónicos basado en reglas, similar a los que se utilizan en el sector privado y el público para clasificar las llamadas telefónicas. Se pueden utilizar enfoques similares para algunos tipos de bots conversacionales.[11]

Figura 2.4: Ejemplo de árbol telefónico basado en reglas

FLUJO SIMPLE BASADO EN REGLAS

Inicio

Frase de entrada

Frase = "contestar la llamada"

NO

Frase = "aceptar la llamada"

NO

Frase = "unirse a la llamada"

NO

SI

SI

SI

Ejecutar "aceptar la llamada"

No fue posible comprender la frase

Fin

Fuente: https://medium.com/botsupply/rule-based-bots-vs-ai-bots-b60cdb786ffa.

[10] En este contexto, la interpretabilidad plantea la pregunta "por qué el sistema hizo eso" mientras que la explicabilidad plantea "cómo hizo eso el sistema" de https://louisabraham.github.io/ai-trust/slides.pdf.

[11] https://chatbotsmagazine.com/which-is-best-for-you-rule-based-bots-or-ai-bots-298b9106c81d.

En algunos casos, los sistemas basados en reglas pueden ser eficaces y responder suficientemente a las necesidades de los usuarios, tanto ciudadanos como funcionarios públicos, proporcionando una respuesta rápida a solicitudes directas sin tener que entrenar al sistema con una gran cantidad de datos. Este método de IA fue especialmente popular durante el segundo auge de la IA.

No obstante, el método tiene varias limitaciones. Requiere una cantidad significativa de conocimientos sobre la organización, el proceso y el contexto para los cuales se desarrollan estos sistemas de IA, a fin de ser capaz de articular explícitamente todas las reglas necesarias para programar la aplicación. Otro nombre para la IA basada en reglas es "sistemas expertos", debido a que se necesitan expertos en estas áreas para determinar qué reglas del tipo SI… ENTONCES se necesitan. En segundo lugar, aunque se cuente con conocimiento, puede ser difícil hacerlo explícito mediante la verbalización y codificación. Los expertos en la materia deben colaborar con los ingenieros y científicos de datos para convertir las reglas comerciales conceptuales y de alto nivel, en programas lógicos que las máquinas puedan leer. Por ejemplo, los especialistas en políticas y datos pueden modelar una organización y sus procesos mientras los programadores crean el código real.

Los sistemas basados en reglas pueden ser fáciles de implementar cuando los procesos descritos son relativamente pequeños y sencillos, e implican un número limitado de acciones y de personas para realizarlos. Sin embargo, programar reglas en los casos en que el número de partes interesadas crece y hay que tener en cuenta cada vez más factores, se vuelve algo tedioso o incluso imposible. Cuando hay que tener en cuenta muchos casos inusuales y excepciones, el proceso de programación de reglas se vuelve complejo. Otro aspecto limitante de los sistemas basados en reglas es que las reglas principales suelen definirse una vez y con frecuencia no evolucionan con el tiempo para reflejar los cambios, las nuevas condiciones y las limitaciones. Lo anterior se debe a que cualquier cambio o mejora de las reglas debe hacerse de forma manual mediante la intervención humana (por ejemplo, modificaciones a la programación). Por lo tanto, el mantenimiento de los sistemas basados en reglas suele representar un verdadero desafío. A medida que el número de enunciados del tipo SI…ENTONCES aumenta y se vuelve más complejo, puede resultar complicado añadir nuevas reglas y enunciados sin provocar contradicciones con las reglas existentes. Conforme pasa el tiempo, se genera una base de conocimientos difícil de manejar cuyos beneficios disminuyen por su propio peso. Por consiguiente, la IA basada en reglas es más adecuada para problemas sencillos que se clasifican en una sola área temática y que es poco probable que cambien frecuentemente.[12] Dichas limitaciones en cuanto a la adaptabilidad y autonomía pueden ser un motivo por el que algunos argumentan que los sistemas basados en reglas ya no deberían considerarse IA.[13]

Si bien estas limitaciones suelen considerarse como debilidades, pueden ser una fortaleza en determinadas situaciones. Las reglas son fáciles de programar, y el Aprendizaje Automático podría considerarse una solución demasiado compleja para algunos problemas simples.[14] Es importante destacar que muchas organizaciones, incluyendo los gobiernos, también

[12] www.tricentis.com/artificial-intelligence-software-testing/ai-approaches-rule-based-testing-vs-learning.

[13] Los elementos de este argumento pueden encontrarse en el debate de expertos del Centro Europeo de Estrategia Política en una sesión de alto nivel sobre "Una estrategia de la Unión Europea para la inteligencia artificial". Véase https://ec.europa.eu/epsc/sites/epsc/files/epsc_-_report_-_hearing_-_a_european_union_strategy_for_artificial_intelligence.pdf para consultar el informe recapitulativo de la sesión.

[14] https://deparkes.co.uk/2017/11/24/machine-learning-vs-rules-systems.

utilizan métodos basados en reglas como un primer paso hacia el mundo de la IA. Al utilizar métodos basados en reglas, dichos gobiernos aprenden algunos principios fundamentales y bloques de información de la IA. Una vez que hayan resuelto algunos problemas fáciles, es más probable que lleguen a comprender las limitaciones de estos métodos y se enfrenten a desafíos que no se pueden resolver con las reglas preprogramadas del tipo SI…ENTONCES. En consecuencia, es posible que busquen técnicas más sofisticadas como el Aprendizaje Automático.

Un ejemplo de este método gradual serían los gobiernos que se centran en la Automatización Robótica de Procesos (RPA, por sus siglas en inglés) para automatizar tareas que pudieran ser arduas al realizarse manualmente pero que constituyen acciones repetidas (Recuadro 2.4). Si bien la implementación de técnicas más sofisticadas puede requerir una transformación más fundamental de los procesos subyacentes, la RPA puede ayudar a automatizar los procesos en su forma actual (Eggers, Schatsky y Viechnicki, 2017).

Recuadro 2.4: Automatización Robótica de Procesos

El surgimiento de la Automatización Robótica de Procesos (RPA) como una tecnología distinta, podría tratarse de sistemas basados en reglas mejor comprendidos a los cuales se les cambió el nombre a automatización. Según el manual de estrategias de RPA de la Administración de Servicios Generales de los Estados Unidos:

La Automatización Robótica de Procesos es una "tecnología de automatización de procesos comerciales que automatiza tareas manuales basadas en reglas en su mayoría, son estructuradas y repetitivas y utilizan software robótico, también denominados bots. Las herramientas de RPA esquematizan un proceso para que un robot lo siga, lo que le permite operar en lugar de un humano."

¿Cómo se relaciona la RPA con el Aprendizaje Automático y otros tipos de IA?

Mientras que el Aprendizaje Automático se centra en el *aprendizaje*, el aspecto central de la RPA es *hacer*. Desde esta perspectiva, la RPA y el Aprendizaje Automático pueden complementarse entre sí. La creciente automatización a través de la RPA de las tareas que tradicionalmente se realizaban de manera manual genera datos que pueden utilizarse para entrenar los algoritmos de Aprendizaje Automático y obtener nuevos conocimientos a partir de los procesos de trabajo.

En conclusión, la RPA puede considerarse como un primer paso para resolver problemas sencillos mediante la implementación de una IA con principios más antiguos, sobre la cual se puede construir una IA más sofisticada y compleja basada en el Aprendizaje Automático.

Fuente: https://tech.gsa.gov/playbooks/rpa; www.capgemini.com/consulting-de/wp-content/uploads/sites/32/2017/08/robotic-process-automation-study.pdf; https://medium.com/@cfb_bots/the-difference-between-robotic-process-automation-and-artificial-intelligence-4a71b4834788.

Hoy en día los métodos basados en reglas, como la RPA, siguen siendo relevantes y los gobiernos los utilizan con frecuencia como componentes en programas más grandes relacionados con la eficiencia, el gobierno digital y la IA. Esto no quiere decir que dichos

gobiernos no estén aplicando también técnicas como el Aprendizaje Automático. Más bien, pueden estar articulando acciones a través de las cuales se aprovechen las ventajas comparativas de las diferentes tecnologías y técnicas.

Recuadro 2.5: RPA en los Estados Unidos

En agosto de 2018, la Casa Blanca emitió una nueva política sobre el cambio de trabajo de bajo valor a trabajo de alto valor. Entre otras cosas, esta política obliga a los organismos gubernamentales de los EE. UU. a que "introduzcan nuevas tecnologías, como [RPA], para reducir las tareas administrativas repetitivas". En la orientación normativa se señala que la RPA puede ayudar a las administraciones públicas a ahorrar tiempo y dinero al automatizar las tareas manuales y rutinarias y mejorar la precisión al reducir el riesgo de error humano.

La administración de los Estados Unidos ha adoptado diversas medidas para impulsar el desarrollo de la RPA, incluyendo el establecimiento de una comunidad de práctica de RPA en todo el gobierno, demostraciones durante ferias tecnológicas públicas y la organización de sesiones informativas, ayuntamientos y debates abiertos entre los organismos para hablar sobre la RPA y ofrecer sesiones de capacitación voluntarias.

Entre los factores identificados que podrían contribuir de forma positiva al desarrollo de la RPA se encuentran la inclusión de los empleados en el proceso de cambio, y un énfasis en la forma en que la RPA puede hacer que los trabajos de la administración pública sean más interesantes y de gran impacto, en lugar de sólo considerar a la RPA como una herramienta para reemplazar los puestos de trabajo.

Un punto clave, de acuerdo con la Directora de Sistemas Informáticos a nivel federal, Suzette Kent, es la importancia de reinvertir los ahorros obtenidos a partir de la RPA en otras formas de inversión en TI, así como ayudar a los empleados a desarrollarse en empleos que impliquen actividades más estratégicas. Se han propuesto programas de actualización profesional para hacer frente a este desafío. Algunas dependencias también están trabajando en formas de rastrear las ganancias en el gasto público derivadas de la modernización de las TI.

Fuente: www.whitehouse.gov/wp-content/uploads/2018/08/M-18-23.pdf; www.fedscoop.com/rpa-savings-federal-agencies-reinvest-suzette-kent.

Si bien los métodos de la IA basados en reglas fueron el tema principal durante muchos años, en general han sido reemplazados por técnicas más sofisticadas que se adaptan mejor a la complejidad, incertidumbre e interconexión crecientes de los problemas modernos. La más sobresaliente es el Aprendizaje Automático.

Aprendizaje Automático: ¿En qué se diferencia?

> *Un funcionario está aprendiendo si mejora su desempeño en tareas futuras después de hacer observaciones sobre el mundo.*
>
> *Russell y Norvig (2016)*

A diferencia de los sistemas basados en reglas, el Aprendizaje Automático se considera un gran salto en la evolución y la inteligencia de la IA. Para algunos, la IA no es Inteligencia Artificial en lo absoluto a menos que aprenda. Argumentan que la codificación estricta de las reglas del tipo SI…ENTONCES que deben seguir las máquinas, no es lo suficientemente inteligente como para calificar como "IA", incluso si los resultados finales pudieran pasar la Prueba de Turing. Dicho debate intelectual está fuera del ámbito de este manual. Sin embargo, el Aprendizaje Automático se ha convertido en la forma dominante de la IA, debido en gran parte al aumento rápido y exponencial de la disponibilidad de datos y la potencia de procesamiento en los últimos años.

Recuadro 2.6: Concepto clave: Aprendizaje Automático

El Aprendizaje Automático es un método en el que las máquinas aprenden a hacer predicciones en nuevas situaciones con base en datos históricos. El Aprendizaje Automático consiste en un conjunto de técnicas que permiten que las máquinas aprendan de manera automatizada, sin instrucciones explícitas de un humano, basándose en patrones e inferencias. Los métodos de Aprendizaje Automático enseñan con frecuencia a las máquinas a alcanzar un resultado, proporcionándoles muchos ejemplos de resultados correctos denominado "entrenamiento". En otro método, los humanos definen un conjunto de reglas generales y normalmente dejan que la máquina aprenda por sí misma a través del proceso de ensayo y error.

Fuente: OECD (2019), *Artificial Intelligence in Society.*

Lo que diferencia a los sistemas de Aprendizaje Automático de sus contrapartes anteriores basados en reglas es su capacidad para *aprender* a través de la experiencia, de forma muy similar a los humanos. Como ya se ha visto con los sistemas basados en reglas, transmitirle a una computadora cómo funciona el mundo a través de reglas del tipo SI…ENTONCES puede resultar muy complejo por varias razones. Para empezar, una computadora no tiene conocimientos previos del mundo. Partiendo de ello, programar una computadora es complejo debido a que es necesario crear una gran cantidad bloques básicos de información para describir las interacciones. Basta con imaginar de qué forma explicaríamos a los niños cómo funciona el sistema bancario: sólo sería posible emplear un vocabulario limitado y se necesitaría tiempo para ayudarles a comprender conceptos comunes difíciles de describir, como el dinero o la economía. En cambio, pensemos en los diversos casos en que los humanos aprenden sin el uso de conocimientos explícitos, como aprender a utilizar un nuevo teléfono celular sin leer el instructivo. Además, algunas actividades manuales son difíciles de enseñar o aprender utilizando sólo instrucciones escritas, por ejemplo, *andar en bicicleta*. Otro ejemplo son las actividades cognitivas, como el aprendizaje de la lengua a edad temprana. Varias teorías enfatizan el papel de la observación, la repetición y el refuerzo

positivo o la retroalimentación negativa para ayudar a los niños pequeños a adquirir la capacidad de hablar antes de que se codifique y formalice el conocimiento.[15]

El método del Aprendizaje Automático se ajusta a esta perspectiva del aprendizaje. En lugar de instruir explícitamente a las computadoras para que sigan las reglas definidas por los humanos, las computadoras se alimentan de experiencias en forma de datos para extraer el conocimiento y las reglas por sí mismas. A las computadoras no se les enseñan nuevos conocimientos, se les enseña a aprender.

El resto del manual se centra en mayor medida en el Aprendizaje Automático y sus propios subconjuntos, ya que este método corresponde al entusiasmo actual y al método predominante de IA. Si bien la IA basada en reglas puede tener fortalezas relativas para determinados tipos de aplicaciones y situaciones, el auge del Aprendizaje Automático impulsado por la explosión de datos ofrece nuevas posibilidades y oportunidades. Se pueden obtener nuevos conocimientos a partir de los datos y se puede lidiar con nuevas situaciones mediante el uso de computadoras, incluso en los casos en que no es posible definir explícitamente las reglas o describir un problema.

Como se mencionó anteriormente, la IA basada en el Aprendizaje Automático exige un mayor nivel de autonomía por parte de las computadoras, las cuales evolucionan con el tiempo a través del aprendizaje. Sin embargo, la voluntad relativa de la IA plantea desafíos éticos en relación con las cuestiones de titularidad, responsabilidad y rendición de cuentas, aspectos que hoy en día son objeto de gran preocupación. En el Capítulo 4 se analizan algunos de ellos.

Aplicación del Aprendizaje Automático

Para seguir avanzando en el Aprendizaje Automático y lograr el impacto deseado es necesario reflexionar en torno a algunas preguntas clave:

- ¿Cómo funcionan los sistemas de Aprendizaje Automático en términos generales?
- ¿Cuáles son las diferentes formas que utilizan las máquinas para aprender y cómo se pueden utilizar para resolver diversos problemas?
- ¿Qué subconjuntos de la IA existen y se benefician del Aprendizaje Automático?
- ¿Cuáles son los riesgos y los sacrificios del Aprendizaje Automático?

Aprendizaje Automático 101: Principios fundamentales

El Aprendizaje Automático es un término general, como la IA. Antes desglosar el concepto de Aprendizaje Automático, es importante identificar el común denominador en estas diferentes técnicas que justifica la idea de *máquinas de aprendizaje*.

Entrenamiento, pruebas y generalización

En términos generales, el proceso de aprendizaje en el Aprendizaje Automático se puede dividir en tres etapas importantes: entrenamiento, pruebas y generalización.

[15] www.khanacademy.org/test-prep/mcat/processing-the-environment/language/a/theories-of-the-early-stages-of-language-acquisition.

Entrenamiento: Durante la fase de entrenamiento, se expone al sistema de IA a los datos de los que *aprende* aplicando modelos estadísticos. La fase de entrenamiento es similar a la de los humanos que recopilan experiencias y aprenden de éstas creando relaciones entre las mismas. Por lo general, durante el entrenamiento sólo se utiliza una parte de todo el conjunto de datos (véase el Recuadro 2.7). En la Tabla 2.2 se proporcionan datos de aprendizaje de muestra que se utilizarán para los ejemplos de esta sección.

Recuadro 2.7: Datos de aprendizaje y de pruebas

En el Aprendizaje Automático, los datos suelen dividirse al azar en dos partes: datos de aprendizaje (normalmente el 80% del conjunto de datos) y datos de pruebas (normalmente el 20% del conjunto de datos).

Ejemplo: Predecir si una persona elegirá utilizar un auto o el transporte público, dependiendo del clima. El conjunto de datos de aprendizaje podría incluir información sobre el clima, como la perspectiva (soleado, nublado, lluvia, etc.), la temperatura (cálida, templada, fría o valores numéricos reales), el viento (sí, no, o valores numéricos para km/h) y la decisión real que un individuo tomó sobre su medio de transporte (auto o transporte público).

Tabla 2.2: Datos de aprendizaje de muestra

Registro	Perspectiva	Temperatura	Viento	Medio de transporte
1	Soleado	Fría	Sí	Auto
2	Soleado	Cálida	No	Transporte público
3	Lluvia	Fría	Sí	Auto
4	Soleado	Cálida	Sí	Transporte público
…	…	…	…	…
200	Lluvia	Templada	No	Auto

Pruebas y validación: Una vez que se entrenó al sistema, puede generar algunos conocimientos sobre los conjuntos de datos. Sin embargo, es importante asegurarse de que el sistema esté correctamente entrenado para resolver el problema que se determinó en un inicio. Para ello, se utiliza otro subconjunto de datos que se separó previamente. La fase de validación se utiliza para afinar y hacer ajustes a los parámetros del modelo a fin de aumentar su rendimiento (más adelante en este capítulo se explica con mayor detalle el rendimiento de la IA).

Ejemplo: Después de haber entrenado el modelo con base en la información contenida en el conjunto de datos de aprendizaje, debe someterse a prueba para ver si es capaz de predecir correctamente si una persona tomará un auto o el transporte público cuando se enfrente a nuevos datos (el subconjunto de datos de prueba). El hecho de que el modelo no sea capaz de hacer las predicciones correctas con

un nivel de rendimiento adecuado es un indicio de que es necesario modificar los parámetros o quizá debería considerarse un modelo o método diferente. Si el modelo arroja resultados positivos en el conjunto de prueba, se puede seguir trabajando para determinar si se pueden hacer mejoras con el conjunto de validación.

Implementación y generalización: Una vez que el sistema ha sido entrenado y se ha sometido a la validación y las pruebas pertinentes, se implementa en un entorno real. Posteriormente, el sistema trabaja con datos nunca antes vistos que se recopilaron en el terreno operativo en tiempo real para ayudar en la toma de decisiones.

Ejemplo: Después de haber entrenado al sistema de IA utilizando un conjunto de datos que incluyen el uso del transporte por parte de un gran número de ciudadanos y el pronóstico del tiempo de los últimos cinco años, se puede implementar para ajustar mejor el suministro de transporte público con base en las predicciones meteorológicas diarias.

Aunque el proceso puede parecer sencillo, está lejos de ser lineal. Es posible que se requieran muchas iteraciones entre el entrenamiento y las pruebas para lograr el ajuste más adecuado del modelo para la tarea en cuestión. Aun así, la implementación del algoritmo final puede representar un reto, ya que las condiciones de la vida real pueden cambiar drásticamente debido a eventos imprevistos. Se podría añadir un cuarto paso en el proceso en el que se incluya nueva información, lo cual podría significar una actualización completa del modelo o un nuevo ciclo de entrenamiento, pruebas e implementación.

Diferentes maneras en que las máquinas pueden aprender

Tal y como se describió, el aprendizaje es el aspecto central del Aprendizaje Automático. Existen cuatro tipos principales de métodos de aprendizaje que pueden utilizarse:

- aprendizaje supervisado
- aprendizaje no supervisado
- aprendizaje por refuerzo
- aprendizaje profundo.

En esta sección se ofrece una breve descripción de cada tipo de aprendizaje, una explicación básica de su funcionamiento y posteriormente se analizan ejemplos y casos para ilustrar la forma en que pueden utilizarse para ayudar a mejorar la política pública, prestar mejores servicios públicos o hacer más eficientes las operaciones internas.

Aprendizaje supervisado: El arte de hacer predicciones

Principios fundamentales

El aprendizaje supervisado es útil particularmente cuando se ha identificado un problema en específico y se cuenta con suficiente información sobre la estructura y el contenido de los datos. Por lo general, el aprendizaje supervisado se asocia con dos tipos de problemas: *regresión* y *clasificación*. En ambos casos, el objetivo del usuario es generar con facilidad predicciones sobre nuevos puntos de datos a partir de observaciones anteriores. La regresión ayuda a predecir el valor numérico de una variable objetivo, mientras que la clasificación

(también denominada categorización) ayuda a predecir la categoría a la que pertenecerá el nuevo punto de datos.

- **Ejemplo de regresión**: un agente inmobiliario desea predecir el precio de una casa utilizando los datos del mercado inmobiliario. Para ello, el agente debe indicar qué datos debe utilizar el sistema. Diferentes factores como el tamaño, la ubicación y el número de habitaciones funcionan como *características de los datos* y el precio es la *variable objetivo* que se desea predecir.
- **Ejemplo de clasificación**: un banco recopila datos históricos sobre sus clientes y los utiliza para establecer los niveles de riesgo. Posteriormente, el banco puede utilizar los datos para evaluar si es probable que un nuevo solicitante de préstamo lo pague o no (por ejemplo, riesgo alto, riesgo medio y riesgo bajo).

De manera general, el aprendizaje supervisado requiere "datos etiquetados", es decir, datos en los que se conoce un resultado o una respuesta final para decisiones anteriores (para obtener más detalles véase el Recuadro 2.2). El término "supervisado" se refiere a la intervención humana necesaria para seleccionar la variable dependiente (es decir, el resultado, la respuesta o la *etiqueta*) de los datos de entrada para guiar el resultado de la IA.

Tabla 2.3: Predecir el uso de los medios de transporte

Registro	Perspectiva	Temperatura	Viento	Medio de transporte
1	Soleado	Fría	Sí	Auto
2	Soleado	Cálida	No	Transporte público
…	…	…	…	…
200	Lluvia	Templada	No	Transporte público

La columna en color verde representa la información de destino que el usuario está interesado en predecir (en este caso, qué medio de transporte se utilizará). La fila en color anaranjado representa las *características* de los conjuntos de datos (información de apoyo que puede utilizarse para hacer predicciones). En este caso, las predicciones se basan en las condiciones meteorológicas utilizando tres características (perspectiva, temperatura y viento).

Aplicaciones en el mundo real

Regresión

- Muchas de las aplicaciones del Aprendizaje Automático en torno a la regresión involucran al sector financiero y tienen como objetivo predecir los precios, ya sea para las bolsas de valores, la vivienda o cualquier otro tipo de activos.[16] Dichos modelos también pueden ser útiles en el sector energético para predecir la demanda, el precio y la producción de energía para optimizar la gestión energética.[17]

[16] https://towardsdatascience.com/predicting-house-prices-with-linear-regression-machine-learning-from-scratch-part-ii-47a0238aeac1.
[17] www.mdpi.com/1996-1073/12/7/1301/pdf.

- La regresión del Aprendizaje Automático también se puede aprovechar en el campo de los servicios de emergencia para prever la demanda de intervenciones. En 2017, los servicios de bomberos y de emergencia de Queensland, en colaboración con el Departamento de Vivienda y Obras Públicas, diversas universidades y empresas de ciencias de datos lanzaron un proyecto que empleaba las técnicas de regresión para predecir las probabilidades diarias de diferentes tipos de peligros como inundaciones, ciclones, incendios y accidentes de tráfico. Posteriormente, este análisis se incorporó en un estudio de posibles escenarios de demanda de servicios y la propuesta de un sistema de respuesta de IA.[18]

Clasificación

- En el sector de telecomunicaciones, la clasificación se utiliza con frecuencia para comprender mejor los motivos por los que un cliente pudiera optar por cancelar su suscripción o conservarla. Este tipo de análisis se denomina *predicción de abandono de clientes* (*churn prediction*) y, en este caso, comprende dos categorías: 1) clientes que se conservaron y 2) clientes que cancelaron.[19]
- Del mismo modo, las empresas pueden utilizar los datos de Recursos Humanos para hacer una serie de predicciones, como, por ejemplo, si un empleado va a renunciar o no (reducción natural del personal) y para comprender mejor los factores que influyen en esta decisión. Por ejemplo, IBM compartió algunos conjuntos de datos de recursos humanos en Kaggle y solicitó ayuda para diseñar modelos que proporcionaran información a partir de variables como la edad, sexo, nivel del puesto, antigüedad en la empresa, antigüedad en el puesto actual y cantidad de años con el gerente actual.[20]
- Otro uso bastante común es en la detección de correos electrónicos no deseados. Además de los inconvenientes, el correo electrónico no deseado es un asunto importante que puede representar una amenaza directa para las empresas y consumir recursos innecesarios. El aprendizaje supervisado se puede utilizar para entrenar un modelo utilizando correos electrónicos que se etiquetaron previamente como no deseados y luego clasificar los nuevos correos entrantes. Hoy en día, la clasificación de correos electrónicos no deseados del Aprendizaje Automático es objeto de mucha investigación.[21]
- La clasificación también puede utilizarse cuando existen más de dos categorías, como en el caso del *análisis de sentimientos*. Por ejemplo, las empresas pueden utilizar el Aprendizaje Automático para clasificar los tweets de acuerdo con su tono positivo o negativo en un nivel general,[22] así como en categorías más definidas (por ejemplo, feliz, triste o enojado).[23]

---

18 Véase el estudio de caso completo en el sitio web del OPSI en https://oecd-opsi.org/innovations/queensland-fire-emergency-services-futures-service-demand-forecasting-model.
19 https://towardsdatascience.com/churn-prediction-770d6cb582a5.
20 https://hackernoon.com/a-machine-learning-approach-to-ibm-employee-attrition-and-performanceb5d87c5e2415; www.kaggle.com/janiobachmann/attrition-in-an-organization-why-workers-quit.
21 https://ieeexplore.ieee.org/abstract/document/5979035.
22 www.businessinsider.fr/us/twitter-facebook-monitoring-2012-11.
23 www.microsoft.com/developerblog/2015/11/29/emotion-detection-and-recognition-from-text-using-deep-learning.

¿Por qué es útil?

Si bien los problemas de regresión y clasificación pueden parecer bastante básicos en teoría (predecir un número con base en diversos datos o predecir una respuesta a una pregunta del tipo sí/no), los casos de uso práctico mencionados indican que muchos problemas comerciales pueden reformularse como situaciones de *regresión* o *clasificación*. Del mismo modo, en el sector público, muchos problemas pueden expresarse como problemas de aprendizaje supervisado.

El uso de estos métodos de aprendizaje supervisado puede ayudar a tomar decisiones más rápidas en varios organismos públicos, así como decisiones que toman en consideración más datos de los que podría procesar un solo humano responsable de casos. Por ejemplo, desde 2007, el Gobierno de Hong Kong ha estado desarrollando un sistema para agilizar el procesamiento de millones de formularios de solicitud que se reciben en el Departamento de Inmigración, incluyendo la aprobación o la negación de visas.[24] En el Reino Unido, el *Behavioural Insights Team* trabajó en un sistema de apoyo para la toma de decisiones a fin de ayudar a detectar a los niños que necesitan atención social especializada. En el sistema se combinó información de referencia, información sobre los niños y notas de los casos para clasificar a los niños en diferentes categorías de riesgo.[25] En ambos casos se puede reducir el volumen de trabajo de los funcionarios públicos, con lo cual se mejorarían las condiciones de trabajo, se promovería un mejor desempeño, se mejoraría la precisión de los procesos y se reducirían las incongruencias entre los responsables de los casos.

El análisis de sentimientos también podría permitir que los gobiernos y los organismos públicos aprovechen las redes sociales para identificar de manera más eficaz las necesidades de los ciudadanos y reaccionar más rápidamente ante ellas, así como para comprender mejor los efectos de anunciar e implementar determinada política.

Se pueden visualizar muchas otras aplicaciones; sin embargo, es necesario que las partes interesadas del sector público se familiaricen más con la clasificación y sepan qué problemas laborales se pueden resolver utilizando este tipo de aprendizaje automático. En el Capítulo 3 se analizan más casos de uso específicos en el sector público.

Aprendizaje no supervisado: cómo obtener más información a partir de sus datos

Principios fundamentales

El propósito del aprendizaje no supervisado es obtener nuevos conocimientos acerca de los datos disponibles. Está estrechamente relacionado con el concepto de *minería de datos* que "se refiere a un conjunto de técnicas utilizadas para extraer patrones de información de los conjuntos de datos" (OCDE, 2015a). En concreto, los algoritmos de aprendizaje no supervisado ayudan a determinar la estructura subyacente que pudiera existir en un conjunto de datos identificando los elementos en común de los diferentes puntos de datos a través de métodos

[24] Para obtener más detalles sobre la automatización del proceso de visado de Hong Kong, véase: www.cs.cityu.edu.hk/~hwchun/AIProjects/stories/km/ebrain.

[25] Para obtener más información sobre el trabajo en la asistencia social para niños, véase: http://38r8om2xjhhl25mw24492dir.wpengine.netdnacdn.com/wp-content/uploads/2017/12/BIT_DATA-SCIENCE_WEB-READY.pdf.

como la agrupación, las reglas de asociación o el análisis de componentes principales (véase más adelante).

El aprendizaje no supervisado requiere un entrenamiento con datos. Sin embargo, a diferencia del aprendizaje supervisado, no se requiere el uso de datos "etiquetados" (datos en los que se indica de forma explícita el resultado o las respuestas finales de acciones pasadas) para entrenar un modelo. En otras palabras, no es necesaria la supervisión humana para indicarle a la máquina específicamente lo que debe buscar (véanse los datos etiquetados del Recuadro 2.2).

Aplicaciones en el mundo real

Agrupación

La agrupación es tratar de identificar grupos con elementos en común, o *racimos*, existentes en determinado conjunto de datos, los cuales quizá no identificaría de inmediato un observador humano. En el caso de los humanos, puede ser complicado descubrir aspectos en común entre los elementos de determinado conjunto y crear grupos a partir de éste debido a la existencia de una gran cantidad de variables.

En el mundo de los negocios, la agrupación se ha implementado en varias áreas. A continuación, se enumeran algunas de las más prometedoras:

- **Segmentación y definición del perfil de los clientes**: Las empresas pueden utilizar algoritmos de agrupación para segmentar a sus clientes con base en su historial de compras o en los datos recopilados por otros medios, como un programa de afiliación o de lealtad. Cuando se comprenden los diferentes segmentos y se identifican dichos grupos de clientes se puede obtener información clave para tomar decisiones comerciales, como diseñar campañas promocionales y planes de comunicación específicos, elegir la ubicación de un nuevo punto de venta o el mejor momento para lanzar un nuevo producto.[26]
- **Detección de fraudes y anomalías**: En el sector financiero y bancario, las técnicas de agrupación se pueden utilizar para agrupar transacciones o clientes a fin de descubrir valores atípicos que no encajan en ningún grupo. Estos resultados pueden ser indicadores de actividades fraudulentas.[27] Otras áreas, como la vigilancia y seguridad, también pueden beneficiarse del análisis de anomalías.

Las aplicaciones de agrupación involucran los siguientes tipos de análisis:

- El **análisis dentro de los grupos (*intra-cluster*)** sirve para identificar los elementos en común dentro de un grupo.
- El **análisis entre grupos (*inter-cluster*)** sirve para identificar las diferencias entre los grupos.
- El **análisis de valores atípicos** sirve para saber por qué un punto no forma parte de un grupo.

[26] https://towardsdatascience.com/unsupervised-learning-a-road-to-customer-segmentation-17fa2ff09d3d.
[27] www.cssf.lu/fileadmin/files/Publications/Rapports_ponctuels/CSSF_White_Paper_Artificial_Intelligence_201218.pdf.

Para consultar un ejemplo de agrupación en el sector público, véase el Anexo A en torno al caso de estudio sobre el uso de la IA en convocatorias abiertas a la toma de decisiones públicas en Bélgica.

Reglas de asociación

Otra aplicación para el Aprendizaje no Supervisado es descubrir reglas y relaciones entre diferentes variables en conjuntos de datos grandes. Esto se conoce como "minería de reglas de asociación". Los algoritmos involucrados tratan de identificar las relaciones entre las diferentes operaciones. Muchas personas encuentran este enfoque técnico en su vida cotidiana. Por ejemplo, un supermercado puede analizar sus cifras de ventas y determinar qué probabilidades hay de que alguien que compra pan compre también leche o cualquier otro producto. En el Recuadro 2.8 se proporcionan detalles sobre la funcionalidad similar "con frecuencia se compran juntos" de Amazon.

Recuadro 2.8: Funcionalidad "con frecuencia se compran juntos" de Amazon

Es posible que cualquier persona que compre en Amazon se encuentre con la funcionalidad de la empresa "con frecuencia se compran juntos". En dicha sección, que se encuentra en la página de cada producto, debajo de los detalles de éste, se muestran otros productos que los clientes también han adquirido. Esta funcionalidad es el resultado de la versión de Amazon de minería de reglas de asociación que denominan "filtrado colaborativo basado en artículos".

La idea general de la minería de reglas de asociación es crear una lista con todos los pares de artículos y descubrir con qué frecuencia los clientes los compran juntos. Existen productos que los clientes nunca compran en par, por lo que hacer esto para cada producto en existencia no es eficiente. Amazon, por el contrario, reduce la lista de productos a pares verificando los carritos de compras actuales de los clientes. Asimismo, puede sacar ventaja de su página de recomendaciones en la que los clientes "pueden filtrar sus recomendaciones por la línea de producto y tema, calificar los productos recomendados, calificar sus compras anteriores y ver por qué se recomiendan los artículos" (Linden et al., 2003).

Fuente: www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf.

Además de utilizarse como instrumento de mercadotecnia, la minería de reglas de asociación podría proporcionar información interesante sobre el sector de la salud para los diagnósticos médicos, al vincular los factores y los síntomas con la probabilidad de desarrollar enfermedades, o ayudar a crear nuevos medicamentos mediante la comprobación de secuencias de proteínas y sus efectos.[28] También se utiliza con frecuencia para comprender los patrones de acción de las personas cuando visitan un sitio web, tomando en cuenta las diferentes páginas que se visitan, el orden en que lo hacen o los diferentes enlaces en los que hacen clic.

[28] www.upgrad.com/blog/association-rule-mining-an-overview-and-its-applications.

Análisis de componentes principales

El análisis de componentes principales (ACP) es otro aspecto útil del aprendizaje no supervisado. El objetivo es reducir la complejidad de un problema al identificar los principales factores que influyen en éste. En el sector financiero y en otras áreas, el ACP se puede utilizar para la gestión de riesgos, ya que ayuda a identificar los riesgos más importantes respecto a la priorización.[29]

¿Por qué es útil?

Aunque cada día se generan y recopilan volúmenes de datos cada vez mayores, en el informe de la OCDE (2015a) sobre la innovación basada en los datos se encontró que “los datos no estructurados son, por mucho, el tipo de datos más frecuente y, por lo tanto, ofrecen el mayor potencial para la analítica de datos en la actualidad”.

En este contexto, los gobiernos y los organismos públicos podrían obtener beneficios significativos a partir de los sistemas de Aprendizaje Automático que utilizan el aprendizaje no supervisado. Al ayudar en la interpretación de grandes cantidades de datos que están disponibles pero que no se utilizan de manera eficaz, a través del aprendizaje no supervisado se pueden convertir en información práctica para tomar decisiones basadas en datos.

El ACP también podría ayudar a los dirigentes del sector público a comprender mejor las necesidades de los ciudadanos con base en sus interacciones con los servicios públicos o sus reacciones en las redes sociales, y servir para identificar a los grupos con comportamientos en común para programas específicos. Además, la agregación de datos basados en la localización o en el sello de tiempo podría revelar nueva información sobre temas como la respuesta ante emergencias, la vigilancia del medio ambiente y la prevención del delito.

En cuanto a otras formas de aprendizaje, un aspecto positivo es que los algoritmos no supervisados no requieren tanta intervención humana para guiar la generación de resultados. El aprendizaje no supervisado también puede considerarse como una etapa preliminar de un análisis más profundo. Por ejemplo, se puede utilizar en conjunto con el aprendizaje supervisado para construir sistemas de predicción sólidos. El aprendizaje no supervisado se puede utilizar *primero* para identificar características interesantes de un conjunto de datos, y *posteriormente* el aprendizaje supervisado se utiliza para clasificar correctamente grupos de datos según la información conocida previamente (véase la siguiente sección en la que se expone un análisis del aprendizaje supervisado).

Aprendizaje por refuerzo

Principios fundamentales

El aprendizaje por refuerzo es un tipo de Aprendizaje Automático que ha ganado popularidad recientemente debido a los avances en el hardware y las capacidades de procesamiento. El aprendizaje por refuerzo funciona al hacer que un agente (computadora) complete una tarea por medio de la interacción con un entorno. Con base en dichas interacciones, el entorno proporcionará una retroalimentación que hará que el agente adapte su comportamiento.

[29] https://ijpam.eu/contents/2017-115-1/12/12.pdf.

En otros términos, el agente *aprende* a través del *ensayo y error*; el entorno penaliza el error y recompensa el éxito. Luego, conforme pasa el tiempo ajusta automáticamente su comportamiento, lo cual da como resultado acciones más refinadas.

Figura 2.5: Cómo funciona el refuerzo



Fuente: OPSI.

Supongamos que una empresa desea crear un automóvil autónomo. El primer paso sería crear una simulación virtual que reproduzca el entorno vial previsto. El segundo paso sería definir un objetivo o tarea (por ejemplo, que el automóvil no se estrelle contra un obstáculo). En función de los parámetros como la velocidad, la aceleración, el frenado, etc., el algoritmo de refuerzo por medio del cual se opera el automóvil, se ejecutará en el entorno virtual simulado y aprenderá la manera en que dichos factores influyen en el tiempo que puede conducir sin chocar. Si el algoritmo realiza una acción que ocasiona que el auto se estrelle rápidamente, recibe una retroalimentación negativa y aprende a no repetir este comportamiento. Si es capaz de conducir más lejos sin estrellarse, recibe una recompensa positiva y se fomenta el comportamiento. A través de este proceso de refuerzo, la máquina aprende cómo las diferentes entradas y acciones influyen en su capacidad para completar su tarea asignada.

Aplicaciones en el mundo real

La robótica es un área en la que el aprendizaje por refuerzo es muy prometedor. Por ejemplo, el fabricante de robots más importante del mundo, la empresa japonesa Fanuc, utiliza el aprendizaje por refuerzo para crear robots autodidactas. Primero, se coloca a los robots en una línea de ensamblaje donde se entrenan a sí mismos a través del aprendizaje por refuerzo para realizar diferentes tareas, como recoger objetos sin que se les enseñe explícitamente cómo hacerlo. El proceso de entrenamiento dura aproximadamente ocho horas.[30]

El aprendizaje por refuerzo con frecuencia también se asocia con juegos de mesa tradicionales como el ajedrez o "Go", un juego muy complejo con gran popularidad en muchas partes

[30] www.technologyreview.com/s/601045/this-factory-robot-learns-a-new-job-overnight.

de Asia oriental. Los algoritmos de aprendizaje por refuerzo también se han utilizado para jugar y competir en torneos de videojuegos.[31] En ambos casos, el aprendizaje por refuerzo ha contribuido a lograr actuaciones sobrehumanas en estos juegos debido a las estrategias poco convencionales que la computadora fue capaz de desarrollar.[32]

¿Por qué es útil?

Este tipo de algoritmos no requiere mucha supervisión humana: una vez que se han establecido los parámetros, el agente aprende de sus propias acciones y errores. El sistema genera sus propios datos de aprendizaje mediante experimentación y no depende de observaciones recopiladas previamente, a diferencia del aprendizaje supervisado. Por ejemplo, se puede diseñar una computadora que aprenda a jugar al ajedrez mediante el aprendizaje por refuerzo, jugando contra sí misma o contra un humano, en lugar de analizar los datos de juegos anteriores. Este enfoque puede permitir que las computadoras descubran la mejor estrategia posible en lugar de simplemente imitar el comportamiento humano.[33] El sector público se podría beneficiar de muchas maneras de este tipo de algoritmos.

El aprendizaje por refuerzo se puede utilizar durante el proceso de diseñar políticas para descubrir nuevas estrategias para lograr un efecto determinado en relación con la política. En 2018, un equipo de investigadores propuso un modelo[34] que combina el aprendizaje por refuerzo y el aprendizaje profundo, que sirviera para "comprender la evasión fiscal de las empresas conservadoras" con el objetivo de diseñar políticas fiscales eficaces.

Desafíos

Aunque el aprendizaje por refuerzo no requiere que un humano intervenga para que el agente aprenda, aún se debe realizar una cantidad considerable de trabajo previo en sentido ascendente para definir adecuadamente el agente, el entorno y la *política*, y determinar qué recompensas y sanciones se aplican. No se trata de una tarea ordinaria por lo que es posible que se requieran conocimientos altamente especializados. Los resultados finales también deben ser evaluados por operadores humanos, ya que es posible que el sistema desarrolle estrategias inadecuadas a pesar de haber cumplido con las limitaciones iniciales.[35] Si bien este tipo de aprendizaje puede ser útil para determinar nuevas estrategias, es factible que se requieran una gran cantidad de pruebas de ensayo y error por parte del sistema y, en consecuencia, tiempo y recursos, antes de que pueda operar plenamente.

Aprendizaje profundo: un subconjunto del Aprendizaje Automático inspirado en la biología

Principios fundamentales

La última área del Aprendizaje Automático es el **aprendizaje profundo**. Al igual que en el caso de los métodos anteriores de Aprendizaje Automático, el aprendizaje profundo sigue

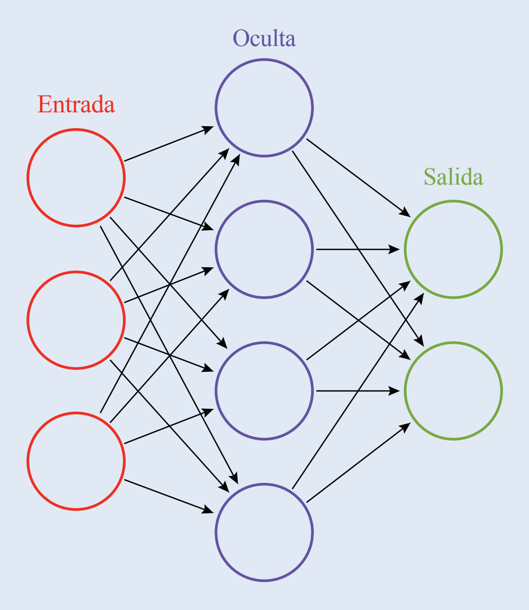
31 https://towardsdatascience.com/the-end-of-open-ai-competitions-ff33c9c69846.
32 www.eurekalert.org/pub_releases/2019-05/aaft-dng052819.php.
33 https://hackernoon.com/reinforcement-learning-and-supervised-learning-a-brief-comparison-1b6d68c45ffa.
34 https://arxiv.org/pdf/1801.09466.pdf.
35 https://vkrakovna.wordpress.com/2018/04/02/specification-gaming-examples-in-ai.

los tres pasos principales: aprendizaje, pruebas y generalización. La principal distinción radica en el diseño de los algoritmos de aprendizaje profundo, el cual se inspira en la biología del cerebro humano. De hecho, el aprendizaje profundo suele analizarse en conjunto con las Redes Neuronales Artificiales (RNA), las cuales se explican en el Recuadro 2.9. La "profundidad" de una Red Neuronal Artificial se relaciona con el número de capas ocultas que posee. Los algoritmos de aprendizaje profundo utilizan las RNA que tienen dos o más capas ocultas. Por ejemplo, se dice que la red ResNet de Microsoft tiene 1 202 capas (Alom et al., 2018).

Recuadro 2.9: Redes Neuronales Artificiales (RNA)

Los científicos estiman que el cerebro humano tiene hasta 100 mil millones de *neuronas*. En concreto, son células nerviosas conectadas entre sí mediante *sinapsis*, las cuales transmiten información enviando impulsos eléctricos de manera bidireccional, durante el proceso de "excitación" o "activación" de las neuronas.

Las Redes Neuronales Artificiales tratan de imitar estos mecanismos y comportamientos utilizando las matemáticas. Los algoritmos de las RNA están diseñados para tener tres componentes principales: una capa de entrada, una capa oculta y una capa de salida.



Cada capa está compuesta por varias *neuronas* o *nodos*. Cada nodo contiene información en forma de un número. Todos los nodos de la capa de entrada están enlazados con nodos de la capa oculta que a su vez están enlazados con nodos de la capa de salida. Estas conexiones son posibles gracias al uso de varias funciones matemáticas (por ejemplo, la función podría consistir únicamente en una suma ponderada de los valores de los nodos de la capa anterior). La transmisión de información de una capa a la siguiente se lleva a cabo mediante otras funciones matemáticas denominadas *funciones de activación*.

Recuadro 2.9: Redes Neuronales Artificiales (RNA) *(Cont.)*

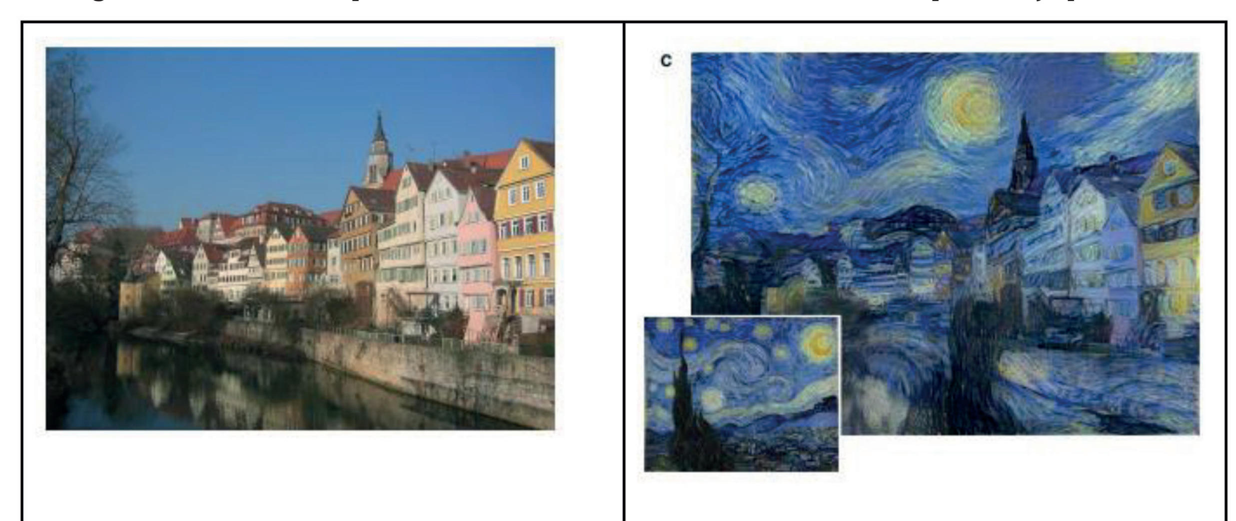
Por ejemplo, en el caso de una RNA utilizada para clasificar imágenes, los nodos de la capa de entrada reciben el valor del color de cada pixel de una imagen; se espera que la capa de salida especifique si la imagen representa un perro o un gato o algo diferente según la aplicación. Durante la fase de entrenamiento, se presentan a la RNA imágenes que ya fueron identificadas como un perro o un gato. Con cada nueva imagen de entrenamiento, la RNA aprende a modificar los coeficientes de sus funciones de activación para producir la respuesta gato/perro esperada.

Fuente: www.ncbi.nlm.nih.gov/pmc/articles/PMC2776484; www.youtube.com/watch?v=aircAruvnKk; https://towardsdatascience.com/intro-to-deep-learning-c025efd92535; https://en.wikipedia.org/wiki/Artificial_neural_network.

Aplicaciones en el mundo real

El aprendizaje profundo se puede utilizar para crear contenido que imite el estilo humano. En la Figura 2.6 se muestran los resultados de extraer características del estilo de pinturas famosas y aplicarlas a una imagen de muestra. MuseNet[36] es una red de aprendizaje profundo a la cual se entrenó con cientos de miles de canciones y la cual aprende las características del estilo de diferentes compositores y músicos, como Frédéric Chopin o los Beatles y es capaz de derivar nuevas piezas musicales e incluso mezclar algunos de los estilos.

Figura 2.6: Creación de pinturas con el estilo de un artista a través del aprendizaje profundo



Fuente: www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Gatys_Image_Style_Transfer_CVPR_2016_paper.pdf.

Uno de los usos más notables del aprendizaje profundo salió a la luz con la aparición de los “ultrafalsos” (*deep fakes*). En 2017, los investigadores de la Universidad de Washington publicaron un documento[37] en el que presentaron un modelo de aprendizaje profundo que aprendía sobre la sincronización entre la forma de la boca y la voz humana a partir de archivos

36 https://openai.com/blog/musenet.
37 http://grail.cs.washington.edu/projects/AudioToObama/siggraph17_obama.pdf.

de audio y video de los discursos del presidente Obama. Posteriormente, el modelo se utilizó para crear un video falso en el que el presidente Obama pronunció un discurso reescrito. En términos más generales, los gigantes de la tecnología como Google o Baidu contribuyen de forma masiva a los sistemas de conversión de texto a voz que producen voces generadas por la IA que leen textos y cada vez son más difíciles de distinguir de la voz humana.[38]

Además de generar nuevas creaciones artísticas, el aprendizaje profundo también puede crear otros algoritmos de aprendizaje profundo y programas informáticos de forma autónoma. Por ejemplo, Google Brain, el equipo que estudia el aprendizaje profundo en Google realizó un experimento[39] en el que hicieron que dos redes neuronales intercambiaran comunicaciones de texto de forma protegida mientras una tercera red trataba de descifrar los mensajes. Los algoritmos de aprendizaje profundo lograron establecer comunicaciones seguras con éxito utilizando su propia técnica de criptografía. Otras herramientas como Neural Complete[40] están utilizando redes neuronales para facilitar la programación de nuevos modelos de aprendizaje profundo.[41]

¿Por qué es útil?

Hoy en día, el aprendizaje profundo es una de las áreas de estudio de la IA más prometedoras. Se puede aplicar a muchos tipos de problemas, incluyendo aquellos que aborda el Aprendizaje Automático estándar, además de problemas más complejos.[42] Algunos expertos en el área, en consecuencia, se refieren a él como un “método de aprendizaje universal” (Alom et al., 2018). Además, los algoritmos del aprendizaje profundo pueden lograr un desempeño impresionante en comparación con las técnicas más tradicionales de Aprendizaje Automático. Por ejemplo, las redes de aprendizaje profundo mantienen el récord de precisión en cuanto a algoritmos utilizados para reconocer dígitos escritos a mano (véanse los estándares de referencia de la MNIST).[43] También se ha atribuido al aprendizaje profundo el mejor desempeño en la detección del cáncer.[44] En general, los algoritmos de aprendizaje profundo al parecer también son los mejores en aprovechar grandes cantidades de datos en comparación con otras formas de Aprendizaje Automático. Un mayor número de datos suele traducirse en un mejor desempeño; sin embargo, el Aprendizaje Automático tiende a alcanzar un límite, mientras que el desempeño del aprendizaje profundo por lo general sigue mejorando cuando se ingresan más datos en la red, siempre que los datos se hayan evaluado para garantizar que son de un nivel de calidad adecuado y se hayan adoptado medidas para mitigar el sesgo.[45]

Desafíos

Entre los principales desafíos del aprendizaje profundo está la incapacidad actual de los seres humanos para comprender plenamente lo que ocurre con exactitud durante el entrenamiento

38 https://arxiv.org/pdf/1609.03499.pdf.
39 https://arstechnica.com/information-technology/2016/10/google-ai-neural-network-cryptography.
40 https://github.com/kootenpv/neural_complete.
41 Para consultar más ejemplos de aplicaciones de aprendizaje profundo, véase www.yaronhadad.com/deep-learning-most-amazing-applications.
42 https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning1b6a99177063.
43 http://yann.lecun.com/exdb/mnist.
44 https://healthitanalytics.com/news/google-deep-learning-tool-99-accurate-at-breast-cancer-detection; https://medium.com/future-today/biomind-artificial-intelligence-that-defeats-doctors-in-tumour-diagnosis-5f8ec97298b2.
45 https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063.

de las redes neuronales (es decir, cómo evolucionan los algoritmos para tomar sus decisiones). Se cree que las diferentes capas de un algoritmo de aprendizaje profundo proporcionan nuevos niveles de abstracción con cada nueva capa que se añade y, por lo tanto, se piensa que las redes con más nodos y más capas suelen ser mejores para resolver problemas más complejos. Por ejemplo, en el caso del reconocimiento de imágenes se supone que una capa puede ser capaz de identificar los bordes de una imagen, mientras que otra puede ser capaz de ensamblar esos bordes para reconocer patrones más complejos como curvas o líneas rectas.

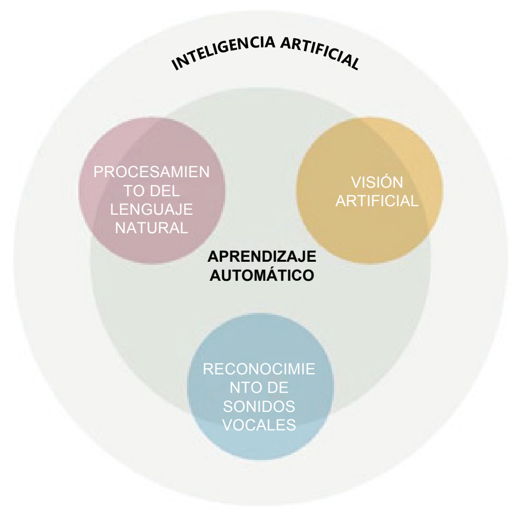
Hasta que se comprenda mejor el funcionamiento interno de los algoritmos de aprendizaje profundo, se les seguirá considerando una *caja negra* en la tecnología. Por lo tanto, su funcionalidad y cualquier resultado que generen representan desafíos en cuanto a *explicabilidad* (véase el Capítulo 4). Otro desafío importante para el aprendizaje profundo es la tensión entre los recursos necesarios para que funcione y el desempeño logrado. Si bien las técnicas de aprendizaje profundo ofrecen la posibilidad de una gran precisión en términos de predicción y pueden encarar problemas más complejos, también requieren computadoras con potentes capacidades de procesamiento y computación, y grandes cantidades de datos para utilizarlos en el entrenamiento.

Otros subcampos de la IA que se benefician del Aprendizaje Automático

Como se mencionó en el Capítulo 1, la IA puede dividirse en muchos subcampos que abordan diferentes tipos de problemas. El auge del Aprendizaje Automático como método distintivo de la IA ha permitido que las comunidades más establecidas reconsideren el tipo de problemas que se pueden resolver, el nivel de desempeño que se puede alcanzar y los recursos necesarios para lograr dichos niveles.

El Aprendizaje Automático también puede utilizarse como un método tecnológico por sí solo, así como en combinación con otros enfoques de la IA (véase la Figura 2.7). Si bien cada uno de estos enfoques podría ejecutarse de una manera basada en reglas, su potencia plena sólo puede lograrse con la introducción del Aprendizaje Automático.

Figura 2.7: Métodos de la IA



Fuente: OCDE basada en https://medium.com/@chethankumargn/artificial-intelligence-definition-types-examples-technologies-962ea75c7b9b.

Al reflexionar sobre estos otros enfoques, puede ser útil pensar en sentidos como la vista, el sonido y el tacto, y en el aprendizaje actuando como una dimensión transversal que vincula y puede ayudar a coordinar los diferentes sentidos. El Recuadro 2.10 presenta el caso de un robot que recoge un objeto, una acción que implica coordinación y la capacidad de tocar y ver. Estas acciones pueden representar subcampos en la IA. Aunque en la sección sobre la *IA estrecha* del Capítulo 1 se describen subcampos adicionales, en esta sección se examinan los que el OPSI ha observado que se utilizan comúnmente o considerado en el sector público hasta la fecha: visión artificial, procesamiento del lenguaje natural y reconocimiento de sonidos vocales.

Recuadro 2.10: Tomar un objeto

"Mire a su alrededor y recoja un objeto con la mano, luego piense en lo que hizo: utilizó sus ojos para explorar el entorno, analizó en dónde se encuentran algunos objetos que pueden tomarse, eligió uno de ellos y determinó una trayectoria para que su mano lo alcanzara, luego movió su mano contrayendo varios músculos en secuencia y logró apretar el objeto con la fuerza adecuada para mantenerlo entre sus dedos".

Las formas en que funcionan los sistemas de IA también se desglosan en métodos similares y adicionales, los cuales en ocasiones pueden realizar una tarea específica por sí solos o combinarse con otros métodos.

Fuente: Elementos del curso en línea de la IA (https://course.elementsofai.com/1/1), OPSI.

Visión artificial

La visión artificial es un subcampo de la IA que se relaciona con la capacidad de ésta para procesar y sintetizar datos visuales (por ejemplo, detectar y clasificar objetos basados en imágenes o videos), con frecuencia a partir de imágenes y archivos de video.

La visión artificial tiene muchas posibles aplicaciones de gran interés para el sector público. En el campo de la medicina, se puede observar una actividad significativa en torno a la detección de enfermedades como el cáncer (véase el Capítulo 3). Los sistemas de reconocimiento de imágenes también funcionan de forma autónoma para escanear las matrículas de los autos, lo que permite el pago de peajes sin dinero en efectivo. Sin embargo, se pueden desarrollar sistemas de visión artificial más sofisticados que apliquen las técnicas de Aprendizaje Automático. Una vez que estas técnicas se combinan, la IA puede aprender, recordar y reconocer imágenes e identificar patrones.[46] El reconocimiento facial es un ejemplo clave de lo anterior. En el sector público, la combinación del Aprendizaje Automático y estas técnicas podría utilizarse para tareas como la identificación de personas, la contratación y el reclutamiento,[47] la vigilancia y la gestión del uso de suelo (véase el Recuadro 3.9 del Capítulo 3). Sin embargo, como se analizó en el Capítulo 4, algunas aplicaciones son controvertidas y su ética es objeto de cuestionamientos.

46 www.quora.com/What-is-the-relation-between-machine-learning-image-processing-and-computer-vision.
47 https://skillroads.com/blog/ai-and-facial-recognition-are-game-changers-for-recruitment.

Además de procesar los datos visuales existentes, algunas aplicaciones implican la construcción de datos visuales, por ejemplo, para crear modelos en 3D a partir de imágenes en 2D.

En el caso de estudio del OPSI sobre la Búsqueda de marcas comerciales australianas que realiza la Oficina de Propiedad Intelectual de Australia (*IP Australia*) (OCDE, 2017b) se puede encontrar un ejemplo de cómo los gobiernos combinan el Aprendizaje Automático y la Visión Artificial. Este servicio permite que los usuarios carguen un logotipo y realicen una búsqueda instantánea en la base de datos de la Oficina de IP de Australia de 400,000 imágenes, la cual posteriormente arroja resultados de la marca con base en la similitud visual mediante el reconocimiento de imágenes.

Procesamiento del Lenguaje Natural

El Procesamiento del Lenguaje Natural (PLN) es un subconjunto de la Inteligencia Artificial que se encarga de la capacidad de las computadoras para manejar e interpretar el lenguaje humano y realizar diversas tareas como la traducción o el análisis de textos.[48] El PLN también se puede combinar con otros subcampos como la visión artificial, por ejemplo, para analizar el texto de documentos escaneados o el texto insertado en imágenes y videos.

El PLN puede tener aplicaciones útiles incluso sin emplear el Aprendizaje Automático. Por ejemplo, se puede utilizar para escanear un gran número de documentos en busca de palabras o frases clave. Sin embargo, los usos más modernos y poderosos del PLN emplean el aprendizaje automático. Estas aplicaciones abarcan desde el filtrado de correos no deseados hasta formas más sofisticadas de análisis de textos, como la traducción en tiempo real y el análisis de sentimientos (véase el Recuadro 2.11). Por ejemplo, un sistema de PLN de aprendizaje automático podría escanear sitios web y publicaciones de las redes sociales para identificar los debates sobre determinados temas (Eggers, Schatsky y Viechnicki, 2017). Además de procesar el lenguaje existente, el PLN, junto con el Aprendizaje Automático, también puede utilizarse para la generación de textos (véase el Recuadro 2.12).

Los organismos públicos ya están utilizando el PLN para prestar servicios más personalizados a los ciudadanos y las empresas con base en interacciones específicas (véase el Recuadro 3.7 del Capítulo 3 sobre los asistentes virtuales en Letonia y Portugal, y el caso de estudio del Anexo A sobre el escenario “bomba en una caja” de Canadá).

Recuadro 2.11: Análisis de sentimientos en las redes sociales en Kenia

En un estudio publicado en 2019, los investigadores Chris Mahony, Eduardo Albrecht y Murat Sensoy trataron de utilizar la IA para conocer la relación entre el discurso en línea y la violencia política en el contexto de Kenia.

Su hipótesis inicial fue que el lenguaje utilizado por figuras influyentes puede estar relacionado con el aumento de la tensión y el riesgo de violencia. Para probar esta

[48] Para conocer más detalles y ejemplos interesantes del PLN, véase la entrada del blog del Gobierno del Reino Unido sobre *Natural Language Processing in Government (Procesamiento del Lenguaje Natural en el Gobierno)* en https://dataingovernment.blog.gov.uk/2019/06/14/natural-language-processing-in-government.

Recuadro 2.11: Análisis de sentimientos en las redes sociales en Kenia *(Cont.)*

hipótesis, los investigadores recopilaron y analizaron datos de Twitter mediante un programa de PLN que analizó las puntuaciones de los sentimientos de los tweets (positivos, negativos o violentos), para rastrear los cambios en las emociones de los actores políticos kenianos influyentes.

El software combinó elementos de PLN y Aprendizaje Automático. Por ejemplo, utilizó un método de "lista de palabras" para calcular las puntuaciones de sentimiento: a cada tweet se le atribuyó una puntuación con base en la frecuencia de las palabras utilizadas y la intensidad asociada a palabras específicas. Posteriormente, se empleó el aprendizaje profundo para convertir los datos no estructurados (un tweet o una entrada de blog) en datos estructurados (una puntuación numérica). Se utilizó un algoritmo de "bosques aleatorios" (*random forest*) para relacionar la puntuación de los sentimientos con las muertes diarias reportadas a través del Proyecto de Datos de Ubicación y Sucesos de Conflictos Armados (*Armed and Conflict Location and Event Data Project*).

Al final, los investigadores descubrieron que su modelo podía predecir con precisión los aumentos y disminuciones en el promedio de víctimas con hasta 150 días de anticipación. El resultado prometedor debe considerarse como un primer paso hacia un sistema de IA que podría anticiparse a los conflictos políticos y ayudar a adoptar medidas para evitar víctimas. Este sistema también ayudaría a comprender mejor la relación entre el lenguaje y la violencia.

Fuente: www.theigc.org/wp-content/uploads/2019/02/Language-and-violence-in-Kenya_Final.pdf.

Recuadro 2.12: OpenAI GPT-2: Sistema de PLN para la generación de texto

En febrero de 2019, la empresa de investigación OpenAI publicó un documento en el cual presentó el modelo GPT-2, un sistema entrenado para predecir automáticamente la siguiente palabra de un texto. La empresa utilizó aproximadamente 40 GB de texto que se extrajo de páginas de Internet para entrenar al modelo. Como referencia, las obras completas de Shakespeare ocupan 500 MB, 80 veces menos que el conjunto de datos de aprendizaje del GPT-2.

El modelo se puede utilizar para generar oraciones y párrafos completos y coherentes con base en indicaciones breves programadas por humanos. Del mismo modo, si al modelo se le ingresa un texto para que lo analice, se le pueden hacer preguntas relacionadas y éste responderá.

Sin embargo, el sistema tiene límites y riesgos. Los desarrolladores se encontraron casos de texto repetitivo, fallas lógicas (por ejemplo, escritos sobre *incendios que ocurren bajo el agua*) y cambios de tema poco naturales. Además, estos sistemas podrían utilizarse para crear reportajes de noticias, entre otros, por lo que la automatización de noticias falsas sea objeto de preocupación.

Recuadro 2.12: OpenAI GPT-2: Sistema de PLN para la generación de texto *(Cont.)*

Dada la posibilidad de que el GPT-2 se utilice con malas intenciones, los creadores publicaron en línea sólo partes del código fuente del modelo. Esta decisión ha planteado algunas cuestiones importantes acerca de los desafíos morales de hacer que este software sea totalmente de código abierto. Si bien se debe fomentar la transparencia y el trabajo colaborativo, emergen cuestiones importantes sobre las responsabilidades individuales ante la ausencia de un marco jurídico claro y la presencia de un alto riesgo de uso indebido y "militarización" de la tecnología.

Aplicaciones

Hoy en día la predicción de texto es una característica común para hacer correcciones o sugerencias de palabras cuando se escribe un mensaje de texto o un correo electrónico. Los servicios de correo electrónico ahora ofrecen respuestas generadas automáticamente a los correos electrónicos con base en el procesamiento de su contenido.

Otras posibles aplicaciones de este tipo de sistema son el procesamiento de un gran volumen de textos y la elaboración de resúmenes instantáneos. También podría permitir a los usuarios realizar consultas en función de un corpus lingüístico. Ambas aplicaciones podrían ser útiles para los encargados de dictar políticas que tratan de tomar decisiones informadas, pero también suscitan interrogantes sobre la fiabilidad y el posible sesgo de estos sistemas que, como se señaló anteriormente, no están exentos de fallas.

Fuente: https://towardsdatascience.com/openais-gpt-2-the-model-the-hype-and-the-controversy-1109f4bfd5e8; https://openai.com/blog/better-language-models; www.bbc.com/future/story/20190812-how-ai-powered-predictive-text-affects-your-brain.

La falta de recursos para entrenar algoritmos en idiomas distintos del inglés (y el predominio del inglés a nivel general) es un reto que se debe considerar al desarrollar el PLN; cuestiones que hay que tener en mente al crear nuevos servicios basados en el PLN. Diferentes países han desarrollado iniciativas para resolver este problema en particular, por ejemplo, el Laboratorio Italiano de Procesamiento del Lenguaje Natural [49] o la inversión de 90 millones de euros del Gobierno español para impulsar la industria del PLN.[50]

Reconocimiento de sonidos vocales

El reconocimiento de sonidos vocales es otra área de la IA que está estrechamente relacionada con el PLN. La diferencia clave es que se centra en el análisis de audio como entrada para reconocer e interpretar el lenguaje hablado, en lugar de texto. También se puede utilizar la misma tecnología para generar sonidos vocales en lugar de analizarlos. En dichos casos, se le puede definir también como tecnología de conversión de *texto a voz* o *de síntesis de sonidos vocales*.

[49] www.italianlp.it.
[50] https://slator.com/demand-drivers/thats-big-spain-pours-100-million-into-language-technology.

Debido a la gran variación del habla humana (por ejemplo, dialectos, acentos), en su mayor parte, la tecnología de reconocimiento de sonidos vocales ya no emplea métodos basados en reglas y opta por el Aprendizaje Automático (Grabianowski, 2006).

En conjunto con el PLN, el reconocimiento de los sonidos vocales podría afectar en gran medida las formas en que las personas interactúan con sus dispositivos electrónicos para acceder a servicios y controlar dispositivos. Se han logrado avances continuos gracias a la combinación del reconocimiento de sonidos vocales con el Aprendizaje Automático, los cuales están contribuyendo a crear interfaces de voz cada vez más sofisticadas que responden mejor al contexto y que interactúan cada vez más como los humanos. Cabe señalar que de esta manera se genera una mayor aceptación y adopción por parte de los consumidores de tecnologías, como los asistentes personales y domésticos que funcionan mediante comandos de voz.[51] En el Recuadro 2.13 se profundiza sobre los posibles usos del reconocimiento de sonidos vocales que son relevantes para el sector público.

Recuadro 2.13: Mayor acceso a diversa información a través del reconocimiento de sonidos vocales

En 2005, la Agencia de Proyectos de Investigación Avanzada para la Defensa (*Defense Advanced Research Projects Agency*, DARPA) comenzó a financiar el programa de Explotación Autónoma Universal de las Lenguas (*Global Autonomous Language Explotation*, GALE) con la intención de desarrollar un sistema que pudiera “transcribir, traducir y resumir automáticamente tanto el texto como los sonidos vocales”. De esta manera, los funcionarios de inteligencia podrían analizar “las noticias transmitidas en el extranjero, los programas de entrevistas, los artículos periodísticos, los blogs, los correos electrónicos y las conversaciones telefónicas” utilizando menos recursos de los que se necesitarían de otro modo. Dicho sistema combina un componente de reconocimiento de sonidos vocales para transcribir los archivos de audio a texto, un módulo de traducción automática para convertir los datos en idiomas distintos al inglés y un elemento de “destilación” que se puede utilizar para realizar consultas a fin de recuperar los datos más relevantes.

IBM Simpler Voice es otro proyecto de reconocimiento de sonidos vocales que puede convertir el texto en imágenes ilustrativas y en breves notas de voz descriptivas. La tecnología fue desarrollada para luchar contra el analfabetismo. Por ejemplo, los adultos con bajo nivel de alfabetización pueden utilizar la aplicación móvil en el supermercado o en las farmacias para escanear diversos productos y recibir información visual y verbal, como instrucciones de uso o advertencias de seguridad. Los defensores de esta solución afirman que el acceso a más información les permitiría a estos sectores de la población, tomar decisiones más saludables y económicas. El uso de esta aplicación podría extenderse a la tramitación de documentos jurídicos y médicos.

Fuente: https://arstechnica.com/information-technology/2006/11/8186; www.speech.sri.com/projects/GALE; www.ibm.com/blogs/think/2017/07/simpler-voice; www.cio.com/article/3403658/how-aiml-is-helping-to-eradicate-poverty.html.

[51] https://medium.com/swlh/the-past-present-and-future-of-speech-recognition-technology-cf13c179aaf.

Rendimiento del Aprendizaje Automático

El entrenamiento y la generalización de un modelo son mecanismos esenciales de los sistemas de Aprendizaje Automático. Sin embargo, las predicciones o descripciones de un conjunto de datos sólo son útiles si son precisas. Del mismo modo, el tiempo de espera para obtener esta información no es favorable. Los investigadores han desarrollado indicadores o *parámetros* con el fin de evaluar el rendimiento de los sistemas de Aprendizaje Automático y poder comparar diversos algoritmos. A continuación, se muestra una lista no exhaustiva de los indicadores más comunes:

Métricas clave para el Aprendizaje Automático

Tiempo y velocidad

La velocidad a la que aprende un algoritmo también puede ser un indicador importante al momento de seleccionar un método particular del Aprendizaje Automático. Diversos factores pueden explicar la rapidez con que se entrena un algoritmo, entre ellos la potencia del procesamiento de la máquina que lo ejecuta, la cantidad de datos de aprendizaje que deben procesarse, el algoritmo específico utilizado y el código empleado para implementarlo. El ejemplo anterior de los robots japoneses autodidactas demuestra que el proceso de aprendizaje puede tardar varias horas antes de que los robots puedan utilizarse por completo en las operaciones. Asimismo, analizar la solución y asegurarse de que produce resultados significativos puede requerir tiempo, dependiendo del tipo de modelo y método utilizado.

Robustez

La robustez es otro parámetro que puede servir como guía para seleccionar algoritmos de Aprendizaje Automático y es un área de investigación muy activa.[52] En términos generales, la robustez se refiere a la capacidad de un modelo para manejar anomalías, puntos atípicos o ruido que pudieran existir en un conjunto de datos, con el objetivo de generar resultados coherentes. Un algoritmo robusto sería aquel que puede distinguir el ruido de la información interesante.

Del mismo modo, los valores atípicos y las anomalías son observaciones que no siguen la tendencia general y pueden ser el resultado de otro fenómeno. Por ejemplo, en el modelo de predicción del medio de transporte, el uso del transporte público puede estar estrechamente relacionado con el clima soleado y las temperaturas cálidas, mientras que el uso del automóvil puede estar estrechamente vinculado con el tiempo lluvioso y con viento. Sin embargo, puede suceder que un usuario decida tomar el transporte público en un día lluvioso y con viento porque el auto se encuentra en mantenimiento. Es posible también que varias personas decidan utilizar su auto debido a obras viales. Todos estos casos pueden mostrarse como valores atípicos o anomalías en el modelo general que sólo se centró en las características meteorológicas como factores de predicción.

Otro ejemplo clásico es un algoritmo de clasificación de imágenes que etiqueta de forma errónea las imágenes al realizar una alteración en los datos de aprendizaje. Por ejemplo, los

52 https://towardsdatascience.com/the-three-pillars-of-robust-machine-learning-specification-testing-robust-training-and-formal-51c1c6192f8.

investigadores publicaron un artículo en el que demostraron cómo un modelo importante de reconocimiento de imágenes de Google clasificaba incorrectamente los camiones de bomberos como autobuses escolares al realizar pequeñas modificaciones en las imágenes, como el hecho de girarlas.[53]

Es fundamental tener en cuenta este último caso al planear el uso de la IA, ya que pone de manifiesto los posibles riesgos de seguridad que pueden surgir al utilizar la IA y las tecnologías basadas en ella. Esta cuestión se analiza más adelante en el capítulo.

Métricas del aprendizaje supervisado

La *matriz de confusión* es una herramienta útil en el contexto de la clasificación binaria (problemas del tipo sí/no) para calcular métricas de rendimiento.

Tabla 2.4: Una matriz de confusión

		Valores reales	
		Positivo	**Negativo**
Valores previstos	**Positivo**	Verdadero positivo (VP)	Falso positivo (FP)
	Negativo	Falso negativo (FN)	Verdadero negativo (VN)

A partir del ejemplo anterior de predicción del uso del automóvil o del transporte público, los diferentes elementos de la matriz de confusión se pueden entender de la siguiente manera:

- **Verdadero positivo (VP)**. El modelo predijo que el ciudadano elegiría el auto (o el transporte público) y en efecto toma el auto (o el transporte público).
- **Verdadero negativo (VN)**. El modelo predijo que el ciudadano no elegiría el transporte público (o el auto) y en efecto no lo hace.
- **Falso positivo (FP)**. El modelo predijo que el ciudadano no elegiría el auto y en realidad sí lo hace.
- **Falso negativo (FN)**. El modelo predijo que el ciudadano elegiría el transporte público y en realidad no lo hace.

La ***precisión*** es la métrica más común y funciona como una expresión de la frecuencia con la que el sistema de IA proporciona una respuesta correcta. A menudo se representa en forma de porcentaje o cociente que indica la proporción de veces que el sistema de IA hizo la predicción correcta (verdadero positivo + verdadero negativo). Un alto porcentaje de precisión indica que el modelo hace sugerencias correctas la mayoría de las veces. Por ejemplo, el modelo utilizado en el ejemplo del medio de transporte tiene una tasa de precisión del 80%, lo que significa que 8 de cada 10 veces el modelo ha evaluado o evaluará correctamente si un individuo eligió su auto o transporte público. La precisión puede intervenir en dos momentos clave diferentes en un sistema de IA. En primer lugar, desempeña un papel importante durante la fase de pruebas y validación, ya que sirve para determinar si un modelo necesita perfeccionarse. En segundo lugar, es importante para vigilar y medir el rendimiento de los modelos que se han implementado en situaciones del mundo real.

[53] www.zdnet.com/article/googles-best-image-recognition-system-flummoxed-by-fakes.

Otra métrica relacionada con la matriz de confusión es la ***sensibilidad***. Ésta mide el cociente de los verdaderos positivos y es de particular importancia para las aplicaciones en los ámbitos financieros y de la salud. Por ejemplo, al predecir una enfermedad, es fundamental pronosticar correctamente si un paciente está en realidad enfermo (verdadero positivo) para poder tratarlo de manera oportuna. Por otro lado, una predicción correcta o incorrecta es menos significativa si el paciente está realmente sano (falso negativo o verdadero negativo). Existen otras métricas de rendimiento.[54]

Cuando se trata de predicciones distintas a las binarias, es decir, de casos en que el algoritmo necesita clasificar los puntos de datos en más de dos categorías diferentes (por ejemplo, sí/no), es necesario adoptar un método ligeramente distinto y más complejo para considerar todos los casos.

Es vital evaluar todas esas métricas cuando se planea implementar sistemas de IA. Dar prioridad a una métrica sobre otra requerirá debates y sacrificios, y dependerá de la aplicación específica de IA que se esté considerando. Si bien en este debate pueden involucrarse varias partes interesadas, la decisión final no puede ser responsabilidad exclusiva de los operadores técnicos y debe representar los intereses y valores de los ciudadanos.

Métricas del aprendizaje no supervisado y por refuerzo

En el caso del aprendizaje no supervisado y por refuerzo, la evaluación del rendimiento de los algoritmos puede ser una tarea más delicada y dependiente del contexto. De hecho, como se señaló, ambos tipos de aprendizaje implican alguna forma de incertidumbre en torno a la información obtenida, lo que puede dificultar que se determine de inmediato si los resultados son correctos o no.

En el caso de los problemas de agrupación, puede resultar útil buscar una gran *similitud dentro de los grupos o cohesión de los grupos (cluster cohesion)* (los elementos del mismo grupo son muy similares), una baja *similitud entre los grupos* (*inter-cluster similarity*) (los elementos de diferentes grupos no son en absoluto similares) y una gran *separación entre los grupos* (*cluster separation*) (el grupo observado es muy distinto de los demás grupos) a fin de comparar los resultados de diferentes métodos de agrupación. Se pueden consultar más detalles técnicos en los recursos de la Universidad estatal de Kent o en Introducción a la recuperación de información (*Introduction to Information Retrieval)* (Manning, Raghavan y Schütze, 2008) sobre el análisis de agrupaciones y diferentes métricas como la *pureza* o el *Índice de Rand* (con el uso del concepto de *precisión* descrito anteriormente).

En el caso del aprendizaje por refuerzo, una forma de comparar los resultados generados por un algoritmo es rastreando la cantidad y la tasa de retroalimentación positiva que un agente recibe a lo largo del tiempo al adoptar una estrategia determinada.[55]

Aprendizaje Automático: Riesgos y desafíos

De acuerdo con la sección anterior, seleccionar un algoritmo y medir su rendimiento es una tarea complicada pero necesaria que requiere mucho esfuerzo y datos de entrada de todas las partes interesadas que participan en el ciclo de vida del sistema de IA. Esta sección ofrece

54 https://en.wikipedia.org/wiki/Confusion_matrix.
55 https://artint.info/html/ArtInt_267.html.

un resumen de los retos técnicos a nivel general que hay que tener en cuenta al implementar un sistema de IA. En el Capítulo 4 se analizan más a fondo los riesgos y desafíos de carácter menos técnico, así como las posibles estrategias para superarlos.

Generalización, subajuste (underfitting) y sobreajuste (overfitting)

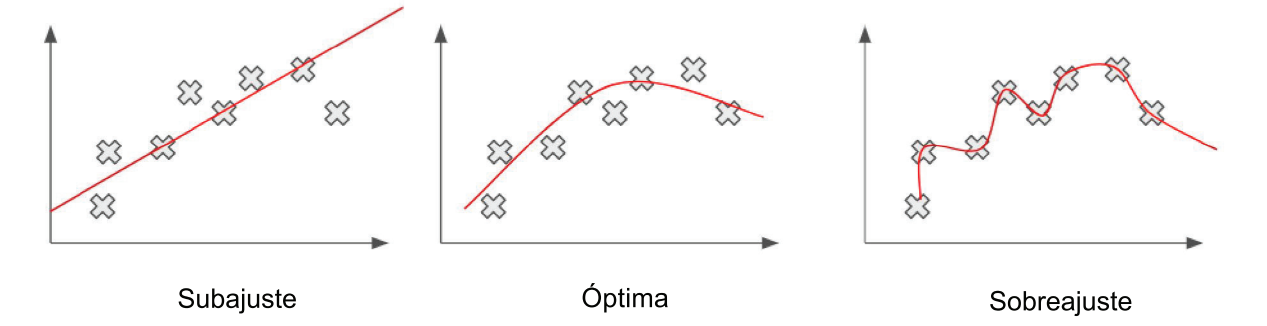
Por lo general, se entiende que el término *generalización* significa cualquier afirmación general formulada respecto a un grupo de personas u objetos a través de la cual determinada realidad para algunos casos se convierte en realidad para todos los casos, y a través de la cual una afirmación que puede ser verdadera a veces se convierta en una que siempre lo es.

En el contexto del Aprendizaje Automático, el término *generalización* se refiere a la "capacidad de un modelo de IA para hacer predicciones correctas sobre nuevos datos nunca antes vistos, en contraposición a los datos utilizados para entrenar el modelo."[56] Como ya se mencionó, en el caso del aprendizaje supervisado y no supervisado, se adquieren nuevos conocimientos con base en datos recopilados previamente y estos nuevos conocimientos se aplican para hacer predicciones. En el aprendizaje por refuerzo, la máquina puede aprender de sus propios errores y el aprendizaje se aplica a nuevas situaciones. Gracias a su rendimiento posiblemente sobrehumano en áreas específicas, es posible que se perciba que los sistemas de Aprendizaje Automático son más inteligentes que los humanos e infalibles. Sin embargo, es importante subrayar que al igual que con cualquier forma de análisis generada por humanos o computadoras: **la correlación no es causalidad y la predicción no es certeza.** Una vez entrenadas, las computadoras pueden hacer predicciones casi instantáneas, sin embargo, éstas deben verificarse. Aunque los algoritmos pueden revelar conexiones en los datos, es vital preguntarse si esas conexiones son coherentes.

El sitio web *Spurious Correlations*,[57] operado por Tyler Vigen, recopila datos de series temporales de varias fuentes y crea gráficos que muestran la correlación entre dos variables. Aunque algunos ejemplos pueden ser graciosos por su obviedad por la diferencia entre los dos conjuntos de datos (un ejemplo relaciona la tasa de divorcio en Maine con el consumo per cápita de margarina, resultando una correlación de más del 99%), otros pueden ser engañosos.

Al analizar la generalización, surgen dos problemas técnicos: *subajuste* y *sobreajuste*.

Figura 2.8: Problemas de generalización: subajuste y sobreajuste



Fuente: https://pythonmachinelearning.pro/a-guide-to-improving-deep-learnings-performance.

56 https://developers.google.com/machine-learning/glossary.
57 www.tylervigen.com/spurious-correlations.

El *subajuste* se refiere a las situaciones en que un sistema de Aprendizaje Automático no puede registrar la información subyacente contenida en los datos. En tales casos, el modelo arroja predicciones incorrectas que son evidentes al examinar las diferentes métricas de rendimiento, como la precisión (véase la sección anterior). El subajuste suele ser el resultado de la aplicación de un modelo inadecuado para el problema en cuestión (es decir, el modelo es demasiado general o demasiado simple con respecto a la complejidad del problema). Si se trata del aprendizaje profundo, aumentar el número de nodos o añadir nuevas capas a la red neuronal puede resultar útil en dichas situaciones.[58] De lo contrario, quizá sea necesario intentar otras técnicas y comparar los resultados.

El *sobreajuste* se refiere a los casos en que el algoritmo es demasiado específico al grado que registra y se centra demasiado en el ruido y las anomalías. Durante la fase de entrenamiento, un modelo de sobreajuste puede alcanzar un alto nivel de precisión y los problemas pueden pasar desapercibidos. Sin embargo, una vez que el modelo entrenado se expone a nuevos datos, la precisión puede disminuir drásticamente. Por ello es necesario realizar las pruebas y validaciones adecuadas antes de implementar un sistema de IA y generalizar los conocimientos adquiridos.

Aunque estos conceptos pudieran parecer ajustes técnicos menores, pueden tener un enorme impacto en ciertas decisiones críticas. La *toma de decisiones basada en la tecnología* puede parecer innovadora; sin embargo, puede crear una falsa sensación de autoridad y confianza. Es responsabilidad de las partes interesadas del sector público asegurarse de que esto no se traduzca en una *mala toma de decisiones basada en predicciones*.

Sesgo, datos y otras cuestiones en materia de seguridad

> *El sesgo no proviene de los algoritmos de la IA, proviene de las personas.*
>
> *Cassie Kozyrkov*[59]

Desde un punto de vista técnico, es importante distinguir entre los diferentes tipos de *sesgo*. Se ha escrito mucho sobre la ética de la IA y en el Capítulo 4 se ofrecen más detalles sobre el problema de los datos y la ética, así como las formas en que los gobiernos pueden encararla.

Cuando se trata de *sesgo de la IA* o *sesgo en algoritmos*, es importante distinguir *el sesgo estadístico*, el cual se refiere sobre todo a un modelo que genera de manera sistemática un error en las predicciones cuando se compara con el resultado esperado. Por ejemplo, en el caso de un modelo de IA de precios de viviendas que predice el valor de un inmueble basándose en los datos de los que dispone pero que constantemente sobrevalora por 1000 euros, el error debe rectificarse antes de implementar el sistema.

Uno de los aspectos que llaman la atención de los responsables de la toma de decisiones respecto al Aprendizaje Automático, es su capacidad para hacer predicciones con base en los activos digitales: los datos. Pero, ¿qué se debe hacer si los datos en sí, y no el modelo, no son los adecuados? De acuerdo con lo que se expuso en la sección anterior sobre “Los datos: el alimento de la IA”, es necesario adoptar medidas cruciales para garantizar la calidad y la representatividad de los datos, y permitir que el modelo no sólo genere predicciones precisas,

[58] www.mikulskibartosz.name/how-to-deal-with-underfitting-and-overfitting-in-deep-learning.
[59] https://towardsdatascience.com/what-is-ai-bias-6606a3bcb814.

sino que también arroje resultados justos para los ciudadanos. Dichas consideraciones se clasifican como otro tipo de sesgo: *sesgo de muestreo* (es decir, sesgo en el proceso de recopilar datos). Un ejemplo de ello podría ser la recopilación de datos sólo para ciertos segmentos de la población.

Más allá del sesgo de muestreo, los gobiernos tendrán que evaluar los datos de los que disponen y determinar si han sido influenciados por factores ya sesgados, y si se puede hacer algo al respecto a fin de evitar que el sesgo se perpetúe cuando los sistemas de IA utilicen los datos. Si los datos reflejan las desigualdades presentes en la sociedad, entonces la aplicación de la IA podría reforzarlas, lo cual a su vez podría distorsionar los desafíos y las preferencias en materia de políticas (Pencheva, Esteve y Mikhaylov, 2018). Según señaló un colaborador durante la consulta pública para este manual, "los datos que se registran no son neutrales, ya que siempre están determinados por el contexto cultural y local. Los datos no caen del cielo; se recopilan, estructuran y almacenan a propósito… Si estos datos se registran en un contexto de discriminación (debido a las prácticas existentes de contratación, por ejemplo), el modelo sólo refleja y magnifica lo que se hizo en el pasado."

Recuadro 2.14: Los sesgos de género y raza en los datos afectan los resultados de los sistemas de IA

En el caso del género, una brecha en los datos desagregados puede representar un desafío para la implementación de las estrategias de IA en la toma de decisiones en el sector público. Según la plataforma técnica y de defensa de los derechos Data2x, "las formas convencionales de datos (encuestas domiciliarias, cuentas de resultados nacionales, registros institucionales, etc.) trabajan en registrar información detallada sobre la vida de las mujeres y las niñas".

Las brechas en los datos sobre la demografía racial también constituyen un problema. En lo que respecta a estudios sobre la salud, la revista *Nature* encontró un "sesgo persistente" en el 96% de los participantes de estudios de todo el genoma (GWAS) de ascendencia europea en 2009, el cual disminuyó ligeramente al 81% en 2016.

Cuando se combinan con los algoritmos de IA y sin una supervisión adecuada, los sesgos en los datos se pueden transferir a la tecnología y por lo tanto generar modelos imprecisos. Por ejemplo, el estudio Gender Shades dirigido por Joy Buolamwini demostró que las mujeres de color eran reconocidas con menor precisión por los sistemas comerciales de reconocimiento facial que los hombres blancos. Los gobiernos de todo el mundo ya están considerando o utilizando dichos sistemas para fortalecer sus políticas de seguridad y reforzar la aplicación de la ley.

En un estudio de Esteva et al. (2017) se analizaron sistemas de aprendizaje profundo con la capacidad de detectar el cáncer de piel con una precisión equivalente a la de los especialistas humanos. Sin embargo, en el estudio se señala que, sin conjuntos de datos debidamente etiquetados, esta precisión no podría medirse en personas con diferentes tipos de piel.

Recuadro 2.14: Los sesgos de género y raza en los datos afectan los resultados de los sistemas de IA *(Cont.)*

Principio de interseccionalidad

La interseccionalidad es un concepto que tiene por objeto explicar las diferentes formas de discriminación que sufre una persona o un grupo de personas, como consecuencia de la superposición de varias identidades sociales, tales como el género, la raza, la orientación sexual y la condición socioeconómica. Es importante aplicar el concepto de interseccionalidad en los conjuntos de datos al diseñar los sistemas y aplicaciones de IA a fin de minimizar los sesgos. La raza y el género son temas particularmente delicados, aunque no los únicos, ya que, si no se tiene en cuenta la condición socioeconómica, los algoritmos podrían reforzar el clasismo, por ejemplo.

La IA puede ayudar a reducir la brecha en los datos

En cambio, la IA se puede utilizar para reducir la brecha en los sesgos mediante el procesamiento de mayores cantidades de datos y la explotación de nuevas fuentes de datos.

Como se muestra en el informe de Data2x, el uso del Aprendizaje Automático ayudó a identificar los síntomas de posibles enfermedades mentales por sexo, con base en un conjunto de datos de medio millón de usuarios de Twitter y aproximadamente 1.5 millones de publicaciones de India, Sudáfrica, Reino Unido y Estados Unidos. Como se señala en el informe del proyecto, “las dificultades a las que se enfrentan las mujeres y los hombres difieren de muchas maneras, y un paso para comprender dichas diferencias es la desagregación por sexo de los datos disponibles”. En este ejemplo se emplearon datos desagregados por sexo en un ámbito en el que con frecuencia se emplea información no desglosada por sexo.

Fuente: https://data2x.org/wp-content/uploads/2019/05/Big-Data-and-the-Well-Being-of-Women-and-Girls_.pdf; www.nature.com/news/genomics-is-failing-on-diversity-1.20759; www.nature.com/articles/nature21056; www.unglobalpulse.org/projects/sex-disaggregation-social-media-posts.

Si bien el sesgo puede ser involuntario, otro riesgo de la IA es la posibilidad de ataques intencionales o manipulación. A pesar de los esfuerzos por curar los datos, las condiciones de la vida real en ocasiones pueden ser desfavorables y los sistemas de IA pueden ser objeto del dolo por parte de personas malintencionadas. En su libro blanco[60] sobre Inteligencia Artificial, la autoridad financiera de Luxemburgo, la *Commission de Surveillance du Secteur Financier* (CSSF) destaca tres tipos de acción: el envenenamiento de datos, el ataque de adversarios y el robo de modelos. El *envenenamiento de datos* se refiere a la manipulación de los datos utilizados para el entrenamiento, lo que provoca que el sistema de IA aprenda información incorrecta, lo cual afecta principalmente los sistemas de IA que se basan en los datos disponibles en línea para actualizar continuamente su entrenamiento. Por ejemplo, las personas pueden generar contenido en las redes sociales para interferir con el funcionamiento de un sistema de IA creado para llevar a cabo análisis de sentimientos, a fin de evitar que haga las predicciones correctas. Como se señaló anteriormente, las

[60] www.cssf.lu/fileadmin/files/Publications/Rapports_ponctuels/CSSF_White_Paper_Artificial_Intelligence_201218.pdf.

imágenes también se pueden alterar de formas no perceptibles para el ojo humano e introducirse en un algoritmo para que éste clasifique de forma errónea las nuevas imágenes. El *ataque de adversarios* es otro tipo de riesgo para la seguridad mediante el cual los atacantes intentan eludir la detección de los sistemas de IA. Por ejemplo, los atacantes pueden tratar de evadir un filtro de correos no deseados basado en la IA enviando diferentes tipos de correo electrónico y sondeando las posibles fallas en el sistema, para después diseñar correos electrónicos que puedan eludir el filtro. Otra preocupación en torno a los sistemas de Aprendizaje Automático es el riesgo de *robo de modelos*. El objetivo de los atacantes en dichos casos es realizar ingeniería inversa y duplicar el algoritmo de la IA para obtener información confidencial. Por ejemplo, los atacantes podrían tratar de recrear los sistemas de predicción del mercado de valores para beneficiarse de las predicciones. Otro de sus intereses podría ser obtener los datos que se utilizaron para entrenar un modelo y saber qué conocimientos podrían adquirir con dicha información.[61]

La seguridad de los sistemas de IA, así como la evaluación de la seguridad de estos sistemas, es un área en constante evolución. El IBM Center for the Business of Government celebró recientemente una serie de mesas redondas con funcionarios del gobierno y expertos en IA, sobre las oportunidades y los desafíos de la IA en el sector público. La seguridad se planteó como un tema crucial pero complicado, y uno de los participantes, Jason Matheny, director del Centro de Seguridad y Tecnología Emergente de la Universidad de Georgetown, señaló que "la ciencia para medir la seguridad de la IA no existe". En el informe posterior del Centro se detallaron algunas medidas para proteger la IA, entre ellas designar personas para vigilarla y reclutar a otras para que ataquen los sistemas de forma deliberada con el fin de identificar las vulnerabilidades (IBM Center for the Business of Government, 2019).[62]

Interpretabilidad, explicabilidad

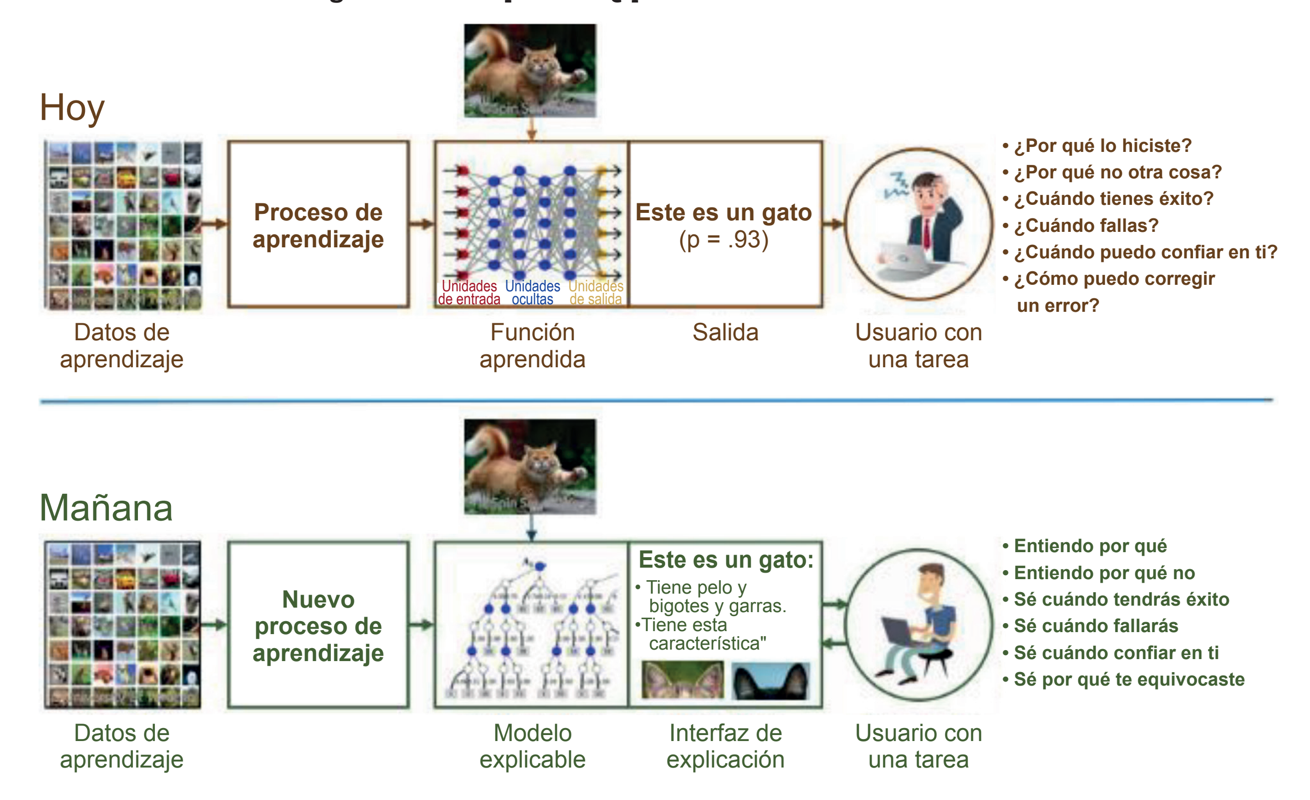
Como ya se mencionó, el aprendizaje profundo es el subconjunto del Aprendizaje Automático más prometedor hoy en día, ya que produce un mejor rendimiento general que cualquier otra técnica. Desafortunadamente, este rendimiento puede verse socavado por la falta de interpretabilidad o explicabilidad. En el caso del aprendizaje profundo, la IA realmente actúa como una caja negra: es capaz de generar resultados; sin embargo, aún no se comprende el proceso por el cual se generan los mismos y los motivos por los cuales el algoritmo toma decisiones específicas.

En algunos casos, la explicabilidad puede significar un problema menor, ya que los resultados en sí son más importantes que el proceso mediante el cual se produjeron (por ejemplo, predecir correctamente si un paciente tiene una enfermedad). Sin embargo, en el caso de los organismos públicos, la explicabilidad es fundamental, ya que las decisiones que se toman con base en la IA deben entenderse y explicarse en su totalidad para efectos de rendición de cuentas y transparencia. Además, las decisiones que toman las partes interesadas públicas pueden influir en gran medida en la vida de los ciudadanos. En el Capítulo 4 se analiza más a fondo este tema en el contexto de las consideraciones y la orientación para el sector público.

[61] En caso de requerir mayor información sobre los ataques contra el aprendizaje automático y las estrategias de defensa visite: https://elie.net/blog/ai/attacks-against-machine-learning-an-overview.
[62] Véase www.businessofgovernment.org/reports para obtener más informes del IBM Center for the Business of Government.

En la Figura 2.9 se muestra el método que la DARPA adoptó para tratar de resolver este problema, modificando el proceso de aprendizaje a través de la inclusión de una etapa que genere un modelo fácil de entender. En la figura se plantea el uso de diagramas de decisiones para vincular los resultados arrojados con las explicaciones.

Figura 2.9: IA explicable: ¿qué estamos tratando de hacer?



Fuente: https://towardsdatascience.com/ai-policy-making-part-4-a-primer-on-fair-and-responsible-ml-and-ai-28f52b32190f.

Subestimación de los humanos

Por último, un problema común al analizar la Inteligencia Artificial es subestimar el papel que desempeñan los seres humanos en el desarrollo no sólo de máquinas de predicción, sino de soluciones completas reales en torno a un problema determinado que pueden o no incluir máquinas de predicción. El caso de Tesla es un buen ejemplo del peligro que representa la excesiva dependencia de la IA y la automatización. En 2017, Tesla anunció su plan para fabricar el nuevo automóvil Model 3 dos años antes de lo previsto, gracias a un método completamente nuevo para fabricar automóviles que implica la "hiperautomatización" (es decir, una línea de ensamblaje automatizada en su totalidad).[63] En 2018, Tesla tuvo que corregir su estimación y la visión subyacente de una fábrica hiperautomatizada, y el CEO, Elon Musk, reconoció que la excesiva automatización fue un error.[64] El ejemplo anterior

[63] https://techcrunch.com/2019/03/05/elon-musk-wasnt-wrong-about-automating-the-model-3-assembly-line-he-was-just-ahead-of-his-time.
[64] www.theguardian.com/technology/2018/apr/16/elon-musk-humans-robots-slow-down-tesla-model-3-production.

es representativo de las situaciones que comúnmente surgen cuando las organizaciones se centran en exceso en soluciones técnicas y no le dan importancia a los conocimientos que poseen los expertos en el área, en particular en lo que respecta a la realización de tareas manuales complejas, o a aquellos que poseen los funcionarios públicos, quienes se enfrentan a problemas específicos en sus labores cotidianas.

Capítulo 3

Prácticas gubernamentales emergentes y el panorama mundial de la IA

Es evidente que la IA está transformando rápidamente muchos aspectos de la vida cotidiana de la gente y que dicha transformación se está acelerando a un ritmo exponencial. El sector público no es inmune, y de hecho se encarga de establecer las prioridades, las inversiones y los reglamentos nacionales cuando se trata de la IA. Lo más relevante de este manual es que los gobiernos también pueden aprovechar el inmenso poder de la IA para innovar y transformar el sector público a fin de redefinir las formas en que diseña y aplica las políticas y presta servicios a su población. Dicha innovación y transformación es fundamental para los gobiernos, ya que se enfrentan a una complejidad y demandas cada vez mayores por parte de sus ciudadanos, residentes y empresas.

La IA puede integrarse en todo el proceso de formulación de políticas y diseño de servicios. A medida que la IA y el Aprendizaje Automático evolucionen, se podrán automatizar más tareas administrativas y de procesos, lo que aumentará la eficiencia del sector público y liberará a los funcionarios públicos para que se concentren en trabajos más significativos. Los gobiernos también serán capaces de comprender mejor y tomar decisiones dentro de sus organizaciones y anticipar las necesidades de los ciudadanos. En caso de hacerlo de manera adecuada, los procesos automatizados pueden ayudar al gobierno a tomar decisiones más justas y precisas que antes.

En este capítulo se analiza cómo los gobiernos de todo el mundo se están adaptando a las nuevas posibilidades y a las nuevas realidades que plantea la IA para transformar el gobierno, y cómo están desarrollando capacidad para anticiparse y prepararse para lo que la IA les depare en el futuro. Aprovecha y se basa en la labor de previsión de la Unidad de Gobierno Digital y Datos Abiertos de la OCDE (*OECD Digital Government and Open Data Unit*) [1]

[1] www.oecd.org/governance/digital-government.

y el Grupo de Trabajo de Altos Funcionarios en Gobierno Digital (*E-Leaders*),[2] así como en la labor que la OCDE ha emprendido para elaborar los Principios de la OCDE sobre Inteligencia Artificial (*OECD Principles of Artificial Intelligence*) y el futuro Observatorio de Políticas en materia de IA de la OCDE.[3]

Estrategias gubernamentales de IA

Como se mencionó anteriormente (Recuadro 1.10), la Recomendación de la OCDE sobre IA hace énfasis en la colaboración internacional como principio clave para el desarrollo satisfactorio de una IA fiable. El compromiso con dicha colaboración se refleja en varias declaraciones recientes. Más recientemente, el G20 adoptó los "Principios de IA del G20",[4] que se derivan directamente de la Recomendación de la OCDE. En 2018, todos los países miembros de la Unión Europea firmaron la Declaración de Cooperación en Inteligencia Artificial (*Declaration of Cooperation on Artificial Intelligence*),[5] por medio de la cual se comprometieron a colaborar para impulsar la capacidad y adopción de la IA en Europa, hacer frente a los problemas socioeconómicos y éticos y garantizar un marco jurídico y ético adecuado. También se comprometieron a que la IA esté a disposición de las administraciones públicas y a que redunde en beneficio de éstas, a compartir las mejores prácticas en la adquisición y el uso de la IA en el gobierno y a implementar prácticas de datos abiertos. El subsiguiente Plan Coordinado Europeo sobre Inteligencia Artificial (*EU Coordinated Plan on Artificial Intelligence*)[6] se basa en la declaración y a través de él se busca "maximizar el impacto de las inversiones a nivel nacional y de la UE, fomentar las sinergias y la cooperación en toda la UE." Asimismo, diez gobiernos firmaron[7] la Declaración sobre Inteligencia Artificial en la región nórdica y báltica *(Declaration on Artificial Intelligence in the Nordic-Baltic Region)*, [8] a través de la cual se comprometen, entre otras cosas, a impulsar el desarrollo de las aptitudes y el acceso a los datos y a formular lineamientos éticos.

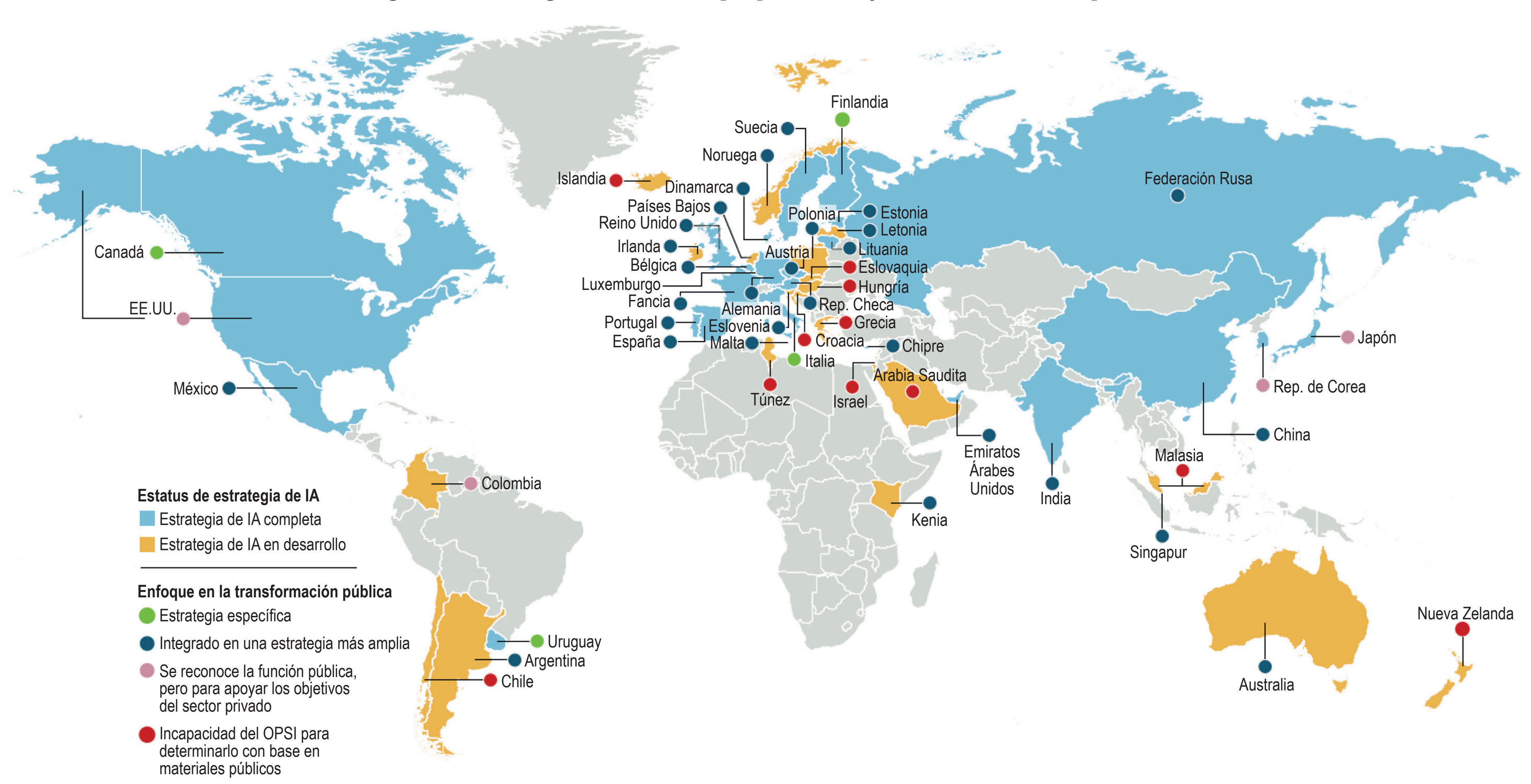
Sin embargo, las estrategias más amplias y detalladas se encuentran a nivel nacional. Muchos países de todo el mundo han adoptado estrategias nacionales de IA o políticas de orientación comparables para establecer visiones y enfoques estratégicos de la IA. Entre ellas se incluyen prioridades y objetivos relacionados con la IA y, en algunos casos, una guía para alcanzarlos. Dichas estrategias pueden ayudar a los países a crear una base común de la IA con miras al éxito, así como a armonizar las capacidades, normas y estructuras de los agentes y ecosistemas pertinentes de la IA. En todo el mundo, al menos 50 países (incluyendo la Unión Europea) han desarrollado o están desarrollando una estrategia nacional de IA (véase la Figura 3.1). Si bien lo anterior implica que una mayoría significativa de países aún no están planeando una estrategia, sí quiere decir que muchos países consideran a la IA actualmente como una prioridad nacional.

Algunos temas comunes emergen al considerar estas estrategias en su conjunto. Casi todos los países se centran (o tienen la intención de centrarse) en catalizar el desarrollo económico

---

[2] www.oecd.org/governance/eleaders.
[3] http://oecd.ai.
[4] www.mofa.go.jp/files/000486596.pdf.
[5] https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence.
[6] https://ec.europa.eu/digital-single-market/en/news/coordinated-plan-artificial-intelligence.
[7] Dinamarca, Estonia, Finlandia, las Islas Feroe, Islandia, Letonia, Lituania, Noruega, Suecia y las Islas Åland.
[8] www.norden.org/sv/node/5059.

Figura 3.1: Estrategias de IA: hasta qué punto incluyen la transformación pública



Fuente: Análisis del OPSI de las estrategias nacionales al 15 de noviembre de 2019 (véase https://oe.cd/aistrategies); Ubaldi, B. et al. (2019), “State of the art in the use of emerging technologies in the public sector”, *OECD Working Papers on Public Governance*, N° 31, OCDE Publishing, París, https://doi.org/10.1787/932780bc-en.

mediante la financiación de la investigación y la I+ D. Por ejemplo, la Unión Europea ha hecho un llamado al sector público y privado para que aumenten las inversiones en IA en al menos 20 millones de euros para finales de 2020, y ha tratado de impulsar los esfuerzos al asignar 1,500 millones de euros en fondos para la investigación. China también se ha comprometido a invertir miles de millones de euros (equivalente) en proyectos de investigación nacionales. En muchos países se están realizando esfuerzos de financiación similares.

La mayoría de las estrategias también incluyen disposiciones para ayudar a garantizar que los sistemas de IA se diseñen e implementen de manera ética, confiable y segura. Por lo general, también incluyen elementos para fortalecer la reserva nacional de talento de la IA, a menudo mediante programas educativos y de capacitación. Lo más importante para este manual es que la mayoría incluye un enfoque específico sobre el uso y las implicaciones de la IA para la innovación y la transformación del sector público.

Componentes del sector público de las estrategias nacionales

De los 50 países (incluyendo la Unión Europea) que ya cuentan con estrategias nacionales de IA o que las están desarrollando, 36 han impulsado ya sea estrategias para transformar el sector público a través de la IA o un enfoque específico integrado en una estrategia más amplia.[9] En cambio, cuatro de las estrategias abordan la importancia del papel del gobierno en la IA, aunque por lo general en el contexto del apoyo a la economía en general. En el caso de varias de las estrategias en desarrollo, el OPSI no pudo determinar si la estrategia final se centraría en el sector público, debido en gran medida a la falta de declaraciones públicas sobre su contenido. El Anexo A contiene un caso de estudio sobre el enfoque de Finlandia respecto a la IA, que incluye una estrategia que abarca la economía en general, así como una estrategia centrada en el ser humano, en específico en lo que respecta a la innovación y la transformación del sector público.

Al igual que en el caso de las estrategias nacionales más amplias, surgen varios temas clave en las estrategias centradas en el sector público. Algunos de ellos son:

- experimentación con la IA en el gobierno y la identificación de proyectos específicos de IA actualmente en curso o que se desarrollarán en un futuro próximo
- colaboración entre sectores, por ejemplo, a través de asociaciones entre el sector público y el privado y facilitada mediante centros y laboratorios de innovación
- fomento de consejos, redes y comunidades intergubernamentales para promover enfoques de sistemas
- automatización de los procesos gubernamentales habituales para mejorar la eficiencia
- uso de la IA para ayudar a orientar la toma de decisiones gubernamentales (por ejemplo, en la evaluación de políticas, la gestión de emergencias y la seguridad pública)
- gestión estratégica, aprovechamiento y apertura de los datos gubernamentales para desarrollar servicios personalizados y anticipatorios, así como para impulsar la IA en el sector privado

[9] En cuanto a las estrategias en desarrollo, esto se basa en declaraciones públicas sobre el contenido previsto de la estrategia pertinente. Los detalles se pueden encontrar en https://oe.cd/aistrategies.

- brindar orientación sobre el uso transparente y ético de la IA en el sector público
- mejoramiento de la capacidad de la administración pública mediante la capacitación, el reclutamiento, las herramientas y la financiación.

Por ejemplo, varios de estos temas generales pueden verse en la estrategia específica del sector público de Italia (véase el Recuadro 3.1).

Recuadro 3.1: La Inteligencia Artificial al servicio de los ciudadanos (Italia)

En marzo de 2018, el Grupo de trabajo de IA de Italia, dirigido por la Agencia para una Italia Digital (AGID, por sus siglas en italiano), publicó el libro blanco Inteligencia artificial al servicio de los ciudadanos (*Artificial Intelligence at the Service of Citizens*). En el documento se analizan los principales problemas relacionados con la implementación de la IA en el sector público, y plantea una serie de recomendaciones sobre la forma en que el gobierno puede superarlos. En particular, las recomendaciones incluyen lo siguiente:

- Establecer y promover una plataforma nacional dedicada al desarrollo de soluciones de IA, en relación con la calidad de los datos, el código y los modelos, las pruebas de los sistemas de IA antes de su lanzamiento y el suministro de recursos informáticos para la experimentación.
- Divulgar los resultados de los algoritmos de IA para facilitar la reproducibilidad, evaluación y verificación.
- Proporcionar recursos abiertos en idioma italiano.
- Desarrollar sistemas de personalización y recomendación adaptables para facilitar los servicios a los ciudadanos en función de sus necesidades específicas.
- Promover la creación de un Centro Nacional de Competencias a fin de impulsar la IA en el sector público y elaborar, entre otras cosas, un manifiesto para el uso de la IA en el sector público.
- Facilitar habilidades promoviendo una certificación de la IA y estableciendo vías de capacitación.
- Proporcionar un plan "Administración Pública 4.0" para impulsar las inversiones públicas en la IA.
- Fomentar la colaboración intersectorial y en toda Europa.
- Establecer un Centro Transdisciplinario sobre la IA a fin de promover el debate y la reflexión en torno a la ética de ésta, y a fin de involucrar a expertos y ciudadanos en las consideraciones necesarias para establecer reglamentos, normas y soluciones. Definir los lineamientos para la IA segura gracias a su diseño y facilitar el intercambio de datos sobre ciberataques en toda Europa.

Fuente: https://ia.italia.it/assets/whitepaper.pdf.

De manera similar a la financiación para mayor I+D, los gobiernos y los organismos internacionales están obteniendo financiación para proyectos en los que participa el sector

público. Por ejemplo, la Unión Europea se ha comprometido a destinar 2.500 mil millones de euros a las asociaciones entre el sector público y el privado,[10] y los gobiernos de Finlandia, Italia, Portugal y Eslovenia declararon que cada uno invertirá más de 10 millones de euros en proyectos del sector público. El documento de la OCDE sobre el *Estado de la Técnica de las Tecnologías Emergentes en el Sector Público (State of the Art on Emerging Technologies in the Public Sector)* (Ubaldi et al., 2019) ofrece un panorama más detallado del apoyo financiero gubernamental para la adopción de tecnologías emergentes en el sector público.

El OPSI ha elaborado un documento complementario a este informe titulado Estrategias de IA y componentes del sector público (AI Strategies & Public Sector Components), en el que se analiza la estrategia nacional de IA completa o en desarrollo de cada país, incluyendo la medida en que cada programa aborda específicamente la innovación y la transformación del sector público. El sitio también incluye enlaces a documentos clave sobre estrategias y políticas. Este recurso puede consultarse en https://oe.cd/aistrategies. Con el tiempo, el OPSI planea integrar este recurso con el[11] próximo depósito de conocimientos del Observatorio de Políticas en materia de IA de la OCDE sobre estrategias nacionales de IA para que los usuarios puedan obtener periódicamente información completa y actualizada sobre éstas. Además, en el informe de la OCDE sobre el *Estado de la Técnica de las Tecnologías Emergentes en el Sector Público* se proporciona información sobre los agentes clave involucrados en la implementación de dichas estrategias (Ubaldi et al., 2019).

Proyectos de la IA con un fin público

Si bien los gobiernos están desarrollando cada vez más estrategias de IA, existe un enorme potencial para que ésta se aplique en todo el sector público a fin de mejorar la forma en que el gobierno se compromete con los ciudadanos y les suministra servicios.

Los estudios indican que algunos de los efectos más inmediatos de la IA en torno al sector público incluirán automatizar tareas sencillas y orientar las decisiones para que el gobierno sea más eficiente y esté mejor informado (Partnership for Public Service/IBM Center for the Business of Government, 2019). El documento de trabajo de la OCDE sobre gobierno digital, *Estado de la Técnica de las Tecnologías Emergentes en el Sector Público*, respalda este hallazgo y demuestra cómo el uso de la IA puede promover las decisiones políticas basadas en datos, lo que conduce a un mejor gobierno. En el documento de trabajo también se identifican las "principales áreas de aplicación" para la transformación de la IA por parte del gobierno: salud, transporte y seguridad (Ubaldi et al., 2019).

A través del trabajo del OPSI también se ha descubierto que la IA es idónea para fomentar relaciones positivas con los ciudadanos y las empresas y para mejorar las funciones regulatorias del sector público. Estudios recientes han demostrado que la IA tiene un importante potencial en múltiples áreas que servirá para alcanzar los Objetivos de Desarrollo Sostenible (ODS) (IDIA, 2019). En esta sección se estudian dichas áreas y se proporciona una serie de ejemplos de proyectos gubernamentales en el mundo real, la cual no es exhaustiva.

[10] http://europa.eu/rapid/press-release_IP-18-3362_en.htm.
[11] http://oecd.ai.

Mejora la eficiencia y la toma de decisiones gubernamentales

En el contexto gubernamental, uno de los beneficios de la IA más importantes y alcanzables en un futuro inmediato es cambiar la forma en que los propios funcionarios públicos realizan su trabajo. La IA tiene el potencial de ayudar al gobierno a pasar de un trabajo de bajo valor a uno de alto valor [12] y a centrarse, en cambio, en las responsabilidades fundamentales "reduciendo o eliminando las tareas repetitivas, revelando nueva información a partir de los datos… y mejorando la capacidad de las dependencias para alcanzar sus misiones" (Partnership for Public Service/IBM Center for the Business of Government, 2019).

El funcionario público promedio dedica hasta el 30% de su tiempo a documentar la información y a otras tareas administrativas básicas (Viechnicki y Eggers, 2017). Al automatizar o evitar incluso una fracción de dichas tareas, los gobiernos podrían ahorrarse una enorme cantidad de dinero, así como reorientar el trabajo de los funcionarios públicos hacia actividades más valiosas, lo que generaría empleos más atractivos (Partnership for Public Service/IBM Center for the Business of Government, 2018; véase el recuadro 3.2).

Recuadro 3.2: Eliminar tareas tediosas en el Departamento del Trabajo (DOL) de los Estados Unidos

Cada año, la Oficina de Estadísticas Laborales del Departamento de Trabajo (DOL, por sus siglas en inglés) se encarga de analizar cientos de miles de encuestas relacionadas con enfermedades y accidentes de trabajo en empresas y organismos del sector público de todo el gobierno. Este análisis es importante tanto para comprender dichos padecimientos como para desarrollar orientaciones que puedan ayudar a prevenirlas en el futuro. Los empleados de oficina deben aprender un complicado sistema de codificación, leer cada informe y codificar varias características, un proceso laborioso y monótono, el cual consume 25,000 horas de trabajo de los empleados cada año.

A partir de 2014, la Oficina comenzó a experimentar con el uso de la IA para codificar las encuestas, comenzando con las respuestas más fáciles y claras. Con el tiempo, el uso de la IA se incrementó y ahora se utiliza en la mitad de las encuestas. La Oficina ha descubierto que en un día la IA puede codificar tanto en un día lo que un empleado capacitado en un mes, y con un mayor nivel de precisión. Los dirigentes de la Oficina consideraron además que era importante comunicar de forma activa a los empleados los beneficios de la IA, enfatizando que su propósito no era reemplazarlos, sino permitirles que se centraran en tareas más complejas y valiosas. La Oficina también impartió sesiones de capacitación a los empleados sobre el Aprendizaje Automático y la forma en que puede añadir valor a su trabajo.

Fuente: Partnership for Public Service/IBM Center for the Business of Government (2018), *The Future Has Begun*, www.businessofgovernment.org/sites/default/files/Using%20Artificial%20Intelligence%20to%20Transform%20Government.pdf.

Como se analizó en el Capítulo 1, un factor clave para el creciente interés en la IA es la gran cantidad de datos de la que se dispone. Sin embargo, los grandes volúmenes de datos involucrados pueden impedir que los gobiernos extraigan conocimientos útiles, fenómeno

[12] www.whitehouse.gov/wp-content/uploads/2018/08/M-18-23.pdf.

al que comúnmente se le denomina "sobrecarga de información". La IA puede ayudar a los gobiernos a superar la sobrecarga de información, obtener nuevos conocimientos y generar predicciones que les ayuden a tomar mejores decisiones políticas. Corea, por ejemplo, está utilizando el Aprendizaje Automático para identificar oportunidades en todos los ministerios gubernamentales a fin de catalizar la innovación en la economía en general (Recuadro 3.3).

Recuadro 3.3: Plataforma para la inversión y la evaluación (PIE) de I+D de Corea

En Corea, la financiación gubernamental para I+D ha crecido de manera constante; sin embargo, esta tendencia no ha contribuido plenamente a la obtención de resultados económicos innovadores. El Ministerio de Ciencia y TIC ha identificado varios problemas clave, entre los cuales se pueden mencionar:

- Los programas de I+D están fragmentados entre 14 ministerios y dependencias diferentes y el intercambio de información es limitado.
- La investigación básica y fundamental no está relacionada con las etapas posteriores de investigación y desarrollo aplicadas y comerciales.
- Las barreras reglamentarias no se consideran adecuadamente en la etapa de desarrollo.
- El ciclo de retroalimentación entre la evaluación y la financiación no suele estar en consonancia.

Para enfrentar estos problemas y hacer que la I+D a nivel nacional sea más sostenible y pueda anticipar los futuros desafíos y oportunidades, el Gobierno de Corea está implementando un nuevo modelo de inversión en innovación: la "PIE de I+D". Este modelo recopila datos de múltiples áreas (por ejemplo, investigación académica, patentes, tendencias tecnológicas a nivel público y privado, información sobre el impacto económico y otro tipo de información sobre el mercado), y posteriormente aplica la analítica de Big Data y el Aprendizaje Automático para evaluar los cambios revolucionarios en el panorama tecnológico, e identificar superposiciones, posibles oportunidades y eslabones faltantes entre los ministerios coreanos, así como entre las partes interesadas del sector privado y del mundo académico.

Se proporciona una plataforma independiente de la PIE de I+D con datos relevantes de diversas áreas de interés estratégico: vehículos autónomos, medicina de precisión, drones de alto rendimiento, disminución de la contaminación atmosférica, granjas inteligentes, redes eléctricas inteligentes, robots inteligentes y ciudades inteligentes. Corea también está tratando de ampliar la PIE de I+D a otras áreas.

Mediante el uso de la PIE de I+D, el gobierno ha encontrado una manera de identificar los eslabones faltantes en las iniciativas de innovación, fomentar la colaboración entre dependencias, universidades y empresas y resolver los problemas sociales. Al comprender mejor el potencial, la viabilidad y los posibles problemas que pudiera plantear el proyecto, el gobierno está en condiciones de tomar decisiones más informadas sobre aquello en lo que se debe invertir y aquello que se debe evitar.

Fuente: https://oecd-opsi.org/innovations/rd-platform-for-investment-and-evaluation-rd-pie.

Atención médica

La IA ya se está utilizando en el área de la salud de diversas maneras, y su potencial respecto a futuras aplicaciones en el sector público es enorme para los países que tienen servicios de salud nacionales. Como se analizó en el documento *Estado de la Técnica* (Ubaldi et al, 2019), las aplicaciones de la IA, en especial aquellas que involucran el Aprendizaje Automático, pueden servir para interpretar los resultados y sugerir diagnósticos, así como para predecir los factores de riesgo y de esta manera introducir medidas preventivas. Asimismo, pueden sugerir tratamientos y ayudar a los médicos a crear planes terapéuticos altamente individualizados. Si se le combina con los conocimientos de los médicos y otros expertos en medicina, la IA puede proporcionar mayor precisión, mayor eficiencia y resultados más positivos en el área de la salud (véase el Recuadro 3.4 y 3.5).

Recuadro 3.4: Iniciativa de Medicina de Precisión (PMI, por sus siglas en inglés) de los Estados Unidos

La PMI es una iniciativa que se lanzó en 2015 a nivel nacional para abandonar el enfoque de una "fórmula invariable" para suministrar los servicios de salud y, en su lugar, adaptar las estrategias de tratamiento y prevención a las características únicas de las personas, como el entorno, el estilo de vida y la biología.

Con la ayuda de la creación de las tecnologías de "secuenciación de ADN de próxima generación" (NGS, por sus siglas en inglés), la medicina de precisión permite la caracterización molecular detallada de trastornos y cánceres mediante la secuenciación rápida del ADN de los pacientes a un costo accesible. Los algoritmos del Aprendizaje Automático pueden analizar con precisión la información secuenciada y utilizar la gran cantidad de datos de los registros médicos de un individuo en beneficio directo del paciente. Esto ayuda a que los médicos tomen mejores decisiones y desarrollen planes terapéuticos más eficaces.

Fuente: www.healthit.gov/topic/scientific-initiatives/precision-medicine; www.oecd.org/education/ceri/GEIS2016-MadelinReport-Full.pdf.

Recuadro 3.5: Detección de cáncer a través del procesamiento de imágenes habilitado por la IA

El cáncer pulmonar es una de las principales causas de muerte relacionadas con el cáncer, y su detección temprana es crucial para el tratamiento de la enfermedad. Los procesos comunes para diagnosticar la enfermedad tienen altas tasas de falsos positivos y falsos negativos. Dichos errores pueden provocar demoras que impiden a los pacientes recibir un tratamiento eficaz.

Google y Northwestern Medicine, un centro médico académico de Chicago, colaboraron para desarrollar un algoritmo de IA de "aprendizaje profundo" para revisar las imágenes utilizadas para diagnosticar el cáncer pulmonar. El algoritmo fue capaz de revisar las ecografías de forma independiente para predecir si indicaban cáncer. Los investigadores compararon las predicciones del sistema de IA con las de los radiólogos con una experiencia significativa en el área. En todos los casos, las predicciones del sistema de IA fueron tan precisas como las de los radiólogos. En algunas situaciones, el sistema de IA superó a los médicos.

Fuente: www.medicalnewstoday.com/articles/325223.php.

Otro ejemplo es el caso de Mongolia, donde se está experimentando una combinación de tecnologías de IA y de cadena de bloques para ayudar a identificar los medicamentos falsificados antes de que lleguen a manos de los consumidores. Este caso se aborda con mayor detalle en el informe del OPSI titulado Adoptar la innovación en el gobierno: tendencias mundiales 2019 (*Embracing Innovation in Government: Global Trends 2019*).[13]

Transporte

Uno de los usos más difundidos de la IA son los vehículos autónomos, como los vehículos de autoconducción que Uber y diversas compañías automovilísticas importantes están probando. Si bien es cierto que el gobierno tiene la función de regular y comprender las implicaciones de estos vehículos, las oportunidades que ofrecen en cuanto a innovación en el sector público son menos evidentes. En lugar de ello, los gobiernos están utilizando la IA para transformar las formas en que predicen y gestionan los flujos de tráfico, y manejan los posibles problemas de seguridad.

Recuadro 3.6: Proyectos gubernamentales de IA para el transporte

Hangzhou, China

La ciudad de Hangzhou, que tiene una población metropolitana de aproximadamente 6 millones de habitantes, se asoció con la empresa tecnológica Alibaba para lanzar el proyecto “Cerebro de la ciudad” (*City Brain*). La iniciativa utiliza cientos de cámaras en toda la ciudad para recopilar datos en tiempo real sobre las condiciones viales. Estos datos legibles por máquina se centralizan y se introducen en un “hub de IA” que toma decisiones que afectan a los semáforos de 128 cruces de la ciudad. El sistema no se limita a supervisar y ajustar el tráfico en función del volumen de los vehículos; también puede tomar decisiones más estratégicas, como identificar y despejar las rutas para las ambulancias en llamadas de emergencia, reduciendo su tiempo de viaje en un 50%.

Singapur

SMRT Corporation, una empresa de transporte público de Singapur, ha trabajado con la empresa privada NEC en un proyecto piloto que utiliza la IA para predecir la probabilidad de que los conductores de autobuses públicos se estrellen en los próximos tres meses. Si los sistemas de IA indicaban una alta probabilidad de accidente para determinado conductor, se les exige que tomen un curso de capacitación. En el proyecto piloto de IA se utilizaron datos históricos de desempeño vial, y dos científicos de datos observaron el comportamiento de los conductores de autobuses a fin de identificar los posibles factores de riesgo.

Portugal

El Ayuntamiento de Lisboa se asoció con el Laboratorio Nacional de Ingeniería Civil (LNEC) y con un socio académico, Instituto Superior Técnico, para establecer sistemas

[13] Véase https://trends.oecd-opsi.org.

Recuadro 3.6: Proyectos gubernamentales de IA para el transporte *(Cont.)*

de IA que permitieran recopilar, tratar, clasificar y utilizar datos sobre movilidad urbana y contexto específico, a fin de trazar mapas y gestionar el flujo del tráfico de manera integral.

Además, Portugal está implementando un proyecto que tiene como objetivo minimizar el tiempo de respuesta de los vehículos de los servicios médicos de emergencia. Se están desarrollando modelos de predicción que pueden anticipar la demanda de servicios combinando los datos históricos existentes y los datos contextuales de varias fuentes (por ejemplo, el clima) para permitir despliegues más estratégicos de los vehículos.

Fuente: https://trends.oecd-opsi.org; https://govinsider.asia/security/five-chinese-smart-cities-leading-way; www.theaustralian.com.au/business/technology/artificial-intelligence-to-predict-accident-risk-of-bus-drivers/news-story/4e7f8e6a4b7ac6e8715966a86284de16; www.fct.pt/apoios/projectos/consulta/vglobal_projecto?idProjecto=154534&idElemConcurso=12346.

Seguridad

La seguridad es una de las principales áreas de interés para los gobiernos que están explorando el uso de la IA. El término abarca tanto la seguridad física como la seguridad informática, y puede cubrir una amplia gama de sectores de los que se encargan los gobiernos, incluyendo la aplicación de la ley, la prevención y recuperación ante desastres y la defensa militar y nacional. En el documento *Estado de la Técnica* se señala, por ejemplo, que “en el ámbito de la vigilancia, los sistemas de visión artificial y de procesamiento del lenguaje natural pueden procesar grandes cantidades de imágenes, textos y sonidos vocales, para detectar en tiempo real posibles amenazas a la seguridad y el orden público “ (Ubaldi et al., 2019).

Como ejemplo de seguridad física, el Departamento de Transporte del Gobierno de Canadá ha puesto a prueba el uso de la IA para realizar una supervisión basada en el riesgo, mediante el escaneo de la información de la carga aérea previa a la carga, para identificar posibles amenazas. En el Anexo A se expone un caso de estudio de su proyecto piloto “bomba en una caja”. Otro ejemplo es el uso del Aprendizaje Automático que hacen los servicios de bomberos y de emergencia de Queensland para pronosticar la probabilidad de grandes peligros (por ejemplo, ciclones e incendios) a fin de ayudar a asignar sus recursos, como se presenta en la Plataforma de Casos de Estudio del OPSI.[14]

La aplicación de la ley es otra área en la que se está incrementando el uso de la IA. El reconocimiento facial se ha utilizado en varias ciudades del mundo para ayudar a localizar a presuntos delincuentes y luchar contra el terrorismo. Sin embargo, esta práctica puede ser muy controvertida, como se analiza en el siguiente capítulo. La Organización Internacional de Policía Criminal (INTERPOL) es una entidad que utiliza el reconocimiento facial y otros tipos de IA para aplicar la ley, y que publicó Inteligencia artificial y robótica en la aplicación

[14] https://oecd-opsi.org/innovations/queensland-fire-emergency-services-futures-service-demand-forecastingmodel.

de la ley (*Artificial Intelligence and Robotics for Law Enforcement*), [15] en el que se analiza el potencial de la IA en la labor policial y se detallan los proyectos que ya están en marcha en el mundo real.

Los gobiernos han sido objeto de grandes ataques cibernéticos en los últimos años. Por ejemplo, la Oficina de Administración de Personal (OPM) de los Estados Unidos fue víctima de un hackeo que dio lugar a la divulgación de información sumamente confidencial sobre más de 21.5 millones de registros, incluyendo información detallada sobre aprobaciones para efectos de seguridad y las huellas dactilares de 5.6 millones de funcionarios públicos.[16] La IA puede ayudar al gobierno a vigilar los problemas de la red y a detectar irregularidades. Países como Tailandia también están utilizando herramientas de ciberseguridad de IA, y otros han publicado guías para su uso, tal como se indica en el Recuadro 3.7.

Recuadro 3.7: La IA en la seguridad cibernética

Tailandia

"Tailandia está utilizando la IA para vigilar el tráfico de la red y realizar análisis de big data para detectar comportamientos sospechosos de los usuarios, por ejemplo, dos inicios de sesión inusuales con los mismos datos de acceso, pero a cientos de kilómetros de distancia"

Reino Unido

El Centro Nacional de Seguridad Cibernética del Reino Unido publicó una guía sobre Herramientas inteligentes de seguridad (*Intelligent Security Tools*), para ayudar a los usuarios a comprender determinados aspectos del uso de las herramientas de seguridad de la IA existentes y para orientar a aquellos que deseen crear unas internas. Proporciona información útil sobre la forma de establecer necesidades, manejar datos, tener en cuenta los recursos disponibles y sacar el máximo provecho de la IA. Plantea una serie de preguntas que ayudan a determinar si una solución específica de IA es un método adecuado para un problema y conjunto de necesidades particulares.

Fuente: https://govinsider.asia/digital-gov/how-thailand-is-using-ai-for-cybersecurity; www.ncsc.gov.uk/collection/intelligent-security-tools.

Relaciones con ciudadanos y empresas

Además de utilizar la Inteligencia Artificial para abordar temas específicos, los gobiernos también están utilizando las aplicaciones de IA de diversas maneras para interactuar con los ciudadanos, los residentes y las empresas. Un tipo popular de IA que se utiliza tanto en el sector público como en el privado, en especial en las primeras etapas de la exploración de la IA por parte de una empresa, son los bots conversacionales. Los bots conversacionales sencillos utilizan un método basado en reglas para interactuar con los ciudadanos con el fin de hacer cosas como dar respuesta a preguntas frecuentes. Las versiones más sofisticadas

[15] www.unicri.it/news/article/Artificial_Intelligence_Robotics_Report.
[16] www.opm.gov/news/releases/2015/09/cyber-statement-923.

utilizan el Aprendizaje Automático para permitir interacciones más complejas y menos concretas. El uso del aprendizaje por refuerzo (véase el Capítulo 2), permite que los bots conversacionales se perfeccionen continuamente para responder mejor a las necesidades de los usuarios (Recuadro 3.8).

Recuadro 3.8: Asistentes virtuales en Letonia y Portugal

Letonia (UNA)

El Registro Mercantil de Letonia creó UNA, un bot conversacional de asistencia virtual que funciona 24/7 y que proporciona respuestas por escrito a las preguntas más frecuentes de los actuales y futuros empresarios letones, incluyendo actualizaciones del estatus de los documentos de registro presentados. Se puede acceder a UNA a través de la página web del Registro Mercantil, así como a través de Facebook Messenger. Es una alternativa a la visita en persona o a la llamada telefónica, y permite a los usuarios recibir respuestas a preguntas en cualquier momento del día.

El gobierno desarrolló UNA en cooperación con un proveedor privado. Los funcionarios del Registro Mercantil consideran que UNA es un catalizador de la gestión de cambio, ya que permite a los funcionarios públicos delegar el trabajo técnico rutinario y centrarse en tareas de mayor valor. Constantemente los empleados le enseñan al sistema de IA preguntas y respuestas adicionales para que éste pueda responder mejor. Desde su lanzamiento en junio de 2018, UNA ha respondido a más de 22,000 preguntas de casi 4,000 usuarios. Además de responder a las preguntas de los clientes, Letonia también está analizando el uso de UNA como herramienta de capacitación para nuevos empleados.

Con base en el éxito de UNA, Letonia está trabajando en el desarrollo de un asistente virtual general para todo el sector público.

Portugal (Sigma)

El 14 de febrero de 2019, el Gobierno de Portugal lanzó ePortugal, el nuevo Portal de servicios públicos, el cual incluía a Sigma, un bot conversacional de asistencia virtual 24/7 que proporciona respuestas por escrito a preguntas frecuentes de los ciudadanos portugueses. Un usuario registrado o no registrado puede acceder a Sigma a través de ePortugal (en cuyo caso adaptará cada vez más sus respuestas a través del PLN). En caso de que Sigma reconozca que su respuesta no es la adecuada, preguntará al usuario si quiere que se le transfiera con un agente humano, y los pondrá en contacto vía telefónica o correo electrónico según la preferencia del usuario. Sigma ofrece una alternativa a la visita en persona o a la llamada telefónica y permite que los usuarios reciban respuestas a las preguntas a cualquier hora del día. Hasta julio de 2019, Sigma había registrado más de 46,250 interacciones.

Fuente: https://oecd-opsi.org/innovations/una-the-first-virtual-assistant-of-public-administration-in-latvia; www.ur.gov.lv/en/about-us/una; www.ur.gov.lv; https://eportugal.gov.pt/en/inicio.

La IA también se puede utilizar para ayudar a los gobiernos a comprender las opiniones y perspectivas de sus ciudadanos a escalas que antes no eran posibles. Por ejemplo, el uso del Procesamiento del Lenguaje Natural para la agrupación y de técnicas de agrupación

(véase el Capítulo 2) permite que los gobiernos obtengan conocimientos valiosos sobre las opiniones de su pueblo. CitizenLab, una organización de la sociedad civil de Bélgica, trabaja con el gobierno en ese sentido (véase el estudio de caso en el Anexo A).

Regulación

La regulación se refiere al conjunto diverso de instrumentos mediante los cuales los gobiernos establecen requisitos para las empresas y los ciudadanos. La regulación incluye todas las leyes, mandatos formales e informales, normas subordinadas, formalidades administrativas y normas emitidas por organismos no gubernamentales o autorregulados a quienes los gobiernos han delegado facultades regulatorias (OCDE, 2018c).[17]

Si bien los reglamentos y otros tipos de formas de ejecución de la ley a menudo están dirigidos a personas y organizaciones ajenas al sector público, la IA brinda una gran oportunidad para aumentar la capacidad de los gobiernos de mejorar su diseño y aplicación (OCDE, 2019c; OCDE 2019d). Por ejemplo:

- Los funcionarios encargados de vigilar el cumplimiento de las leyes podrían utilizar los numerosos datos de los que disponen, y aplicar las herramientas de aprendizaje automático que les ayuden a predecir los sectores a los que deben dirigir sus esfuerzos. Dichas herramientas podrían utilizarse para identificar áreas en las cuales deben enfocarse y para saber a quién es necesario investigar e inspeccionar.
- El Aprendizaje Automático se puede utilizar para predecir de manera más eficaz el resultado de un posible litigio, garantizando una mayor coherencia entre los dictámenes de los tribunales y los de los funcionarios encargados de vigilar el cumplimiento de las normas.

De esta manera las autoridades pueden optimizar sus operaciones al hacer que los recursos abandonen las actividades no productivas, como investigar empresas que es probable que cumplan con la ley, o iniciar litigios que probablemente sean infructuosos, y se dediquen a actividades que les permitan alcanzar sus objetivos en materia de regulación. En el Recuadro 3.9 se analizan ejemplos del uso de la IA para mejorar las funciones del sector público en materia de aplicación de la ley.

Recuadro 3.9: La Inteligencia Artificial (IA) en las funciones regulatorias

Comisión Australiana de Valores e Inversiones (ASIC, por sus siglas en inglés)

El mayor uso de las herramientas regulatorias de última generación (por ejemplo, la IA, la analítica de datos y las ciencias del comportamiento) es una de las estrategias clave del Plan empresarial 2019-2023 de la ASIC. Como parte de este plan, la organización tiene la intención de establecer “un depósito de datos y el uso de técnicas de inteligencia artificial pertinentes, como el aprendizaje automático y soluciones analíticas de texto y voz”.

[17] La Dirección de Gobernanza Pública de la OCDE y su División de Políticas Regulatorias trabajan para ayudar a los gobiernos a cumplir sus misiones mediante el uso de reglamentos, leyes y otros instrumentos a fin de obtener mejores resultados a nivel social y económico y mejorar la vida de los ciudadanos y las empresas. Su trabajo se puede consultar en http://oecd.org/gov/regulatory-policy.

Recuadro 3.9: La Inteligencia Artificial (IA) en las funciones regulatorias *(Cont.)*

En los últimos años, la ASIC ha explorado la IA en varias áreas, incluyendo:

- Explorar el uso del Procesamiento del Lenguaje Natural para detectar anuncios engañosos en Internet.
- Vigilar la actividad comercial de los mercados financieros.
- Detectar problemas de divulgación en las declaraciones de asesoría financiera.

Para financiar estos experimentos, la ASIC ha destinado 6 millones de dólares australianos al estudio de la tecnología de regulación (RegTech) financiera.

Monitoreo de la contaminación (Israel)

Israel ha implementado sistemas de monitoreo en línea de las emisiones contaminantes de las instalaciones industriales. Actualmente se está explorando el potencial de las técnicas de aprendizaje automático en conjunto con el monitoreo en línea de contaminantes para predecir y, en última instancia, prevenir futuros episodios de contaminación (Laster, 2018).

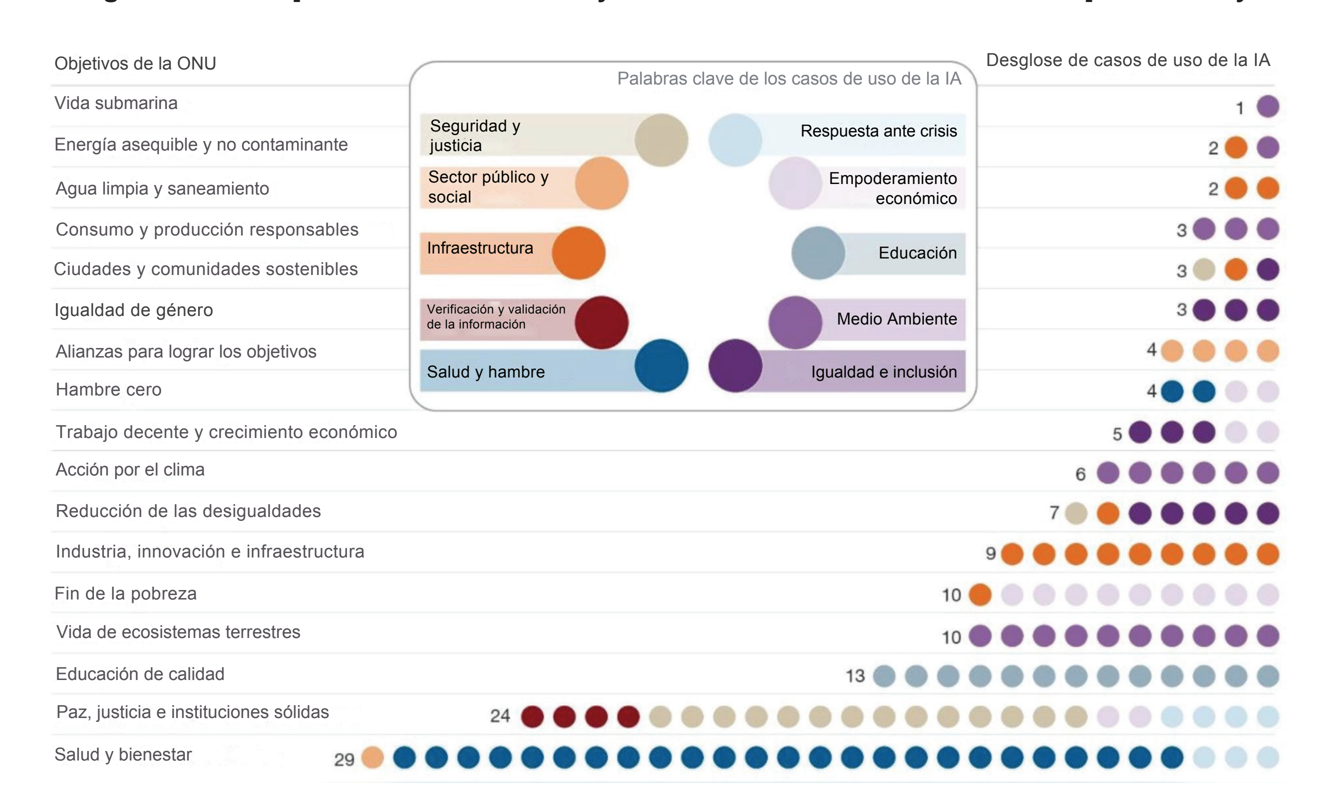
Fuente: https://download.asic.gov.au/media/5248811/corporate-plan-2019-23-published-28-august-2019.pdf; www.afr.com/companies/financial-services/asic-to-use-ai-to-target-misleading-advertising-20190303-h1bx8f, www.innovationaus.com/2018/09/ASIC-puts-6m-to-new-innovation; https://oe.cd/il/techforenforcement.

Objetivos de Desarrollo Sostenible (ODS)

Con la adopción de la Agenda 2030 para el Desarrollo Sostenible, las naciones de todo el mundo se comprometieron a una serie de objetivos y metas universales, integrales y transformadores, denominados Objetivos de Desarrollo Sostenible (ODS). Los 17 objetivos y las 169 metas representan una responsabilidad colectiva y una visión compartida respecto al mundo. Los gobiernos están trabajando para alcanzarlos en el 2030 y muchos están examinando el potencial de la IA para ello.

En los estudios del Instituto Mundial McKinsey (MGI, 2018) se ha identificado un conjunto no exhaustivo de aproximadamente 160 casos que demuestran cómo se puede utilizar la IA “en beneficio de la sociedad sin fines comerciales”. De ellos, 135 coinciden con alguno de los 17 ODS (véase la Figura 3.2). Estos casos suelen adoptar la forma de iniciativas del sector privado o de asociaciones entre el sector privado, el sector público y/o la sociedad civil. Cabe mencionar que el estudio se centra en los ODS como “salud y bienestar”, y “paz, justicia e instituciones sólidas”, pero hace poco énfasis en objetivos como “la vida submarina”, “energía asequible y no contaminante” y “agua limpia y saneamiento”.

Figura 3.2: Correspondencia entre los ODS y los casos de uso de la IA identificados por McKinsey



Fuente: www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Applying%20artificial%20intelligence%20for%20social%20good/MGI-Applying-AI-for-social-good-Discussion-paper-Dec-2018.ashx.

En el sector público, la OCDE se ha dado cuenta de que los gobiernos están tratando de orientar el uso de la IA hacia la preservación del medio ambiente, el capital natural y la resiliencia climática (Ubaldi et al., 2019). Estos objetivos corresponden a algunos de los ODS, como: Agua limpia y saneamiento (ODS 6), Energía asequible y no contaminante (ODS 7), Producción y consumo responsables (ODS 12), Acción por el clima (ODS 13), Vida submarina (ODS 14) y Vida de ecosistemas terrestres (ODS 15). En su informe La Inteligencia Artificial al servicio de la Tierra (*Harnessing Artificial Intelligence for the Earth*),[18] el Foro Económico Mundial (FEM) analiza las maneras en que la IA puede ayudar a encarar los desafíos ambientales. En el Recuadro 3.10 se incluyen algunos ejemplos del sector público.

Recuadro 3.10: Proyectos de la IA que son compatibles con los ODS ambientales

El Aprendizaje Automático en la cartografía

En Australia, el Departamento de Medio Ambiente y Ciencia del Gobierno de Queensland, ha adoptado el Aprendizaje Automático para mapear y clasificar de manera automática las características del uso de suelo de las imágenes satelitales. Identificar los diversos

[18] www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf.

Recuadro 3.10: Proyectos de la IA que son compatibles con los ODS ambientales *(Cont.)*

usos del suelo (por ejemplo, agricultura o vivienda) es crucial para la conservación de la biodiversidad, el monitoreo de desastres naturales y la preparación y respuesta en materia de bioseguridad ante brotes epidémicos. También puede ser útil para proporcionar un análisis casi en tiempo real de los posibles cultivos afectados por grandes desastres como ciclones tropicales e inundaciones. El proceso tiene una precisión del 97%.

Si se utilizan los métodos manuales tradicionales, la cartografía del uso de suelo para todo el estado toma años; por el contrario, el mismo proceso toma sólo seis semanas con la nueva tecnología basada en la IA.

Predicción del consumo de energía

Un centro de investigación del Reino Unido aprovechó los datos de los medidores digitales de electricidad para desarrollar un modelo de IA no supervisado que pudiera predecir qué tipos de dispositivos es probable que se utilicen y cuándo, ello utilizando técnicas de agrupación de Aprendizaje Automático, lo que le permitió predecir los patrones de consumo de energía. Esta información permite a los servicios públicos predecir las futuras necesidades de energía y permite a los residentes calentar sus casas de una manera más inteligente, por ejemplo, apagando automáticamente la calefacción cuando sea probable que no estén en casa. La optimización del consumo de energía puede reducir tanto los precios como el desperdicio de energía.

Fuente: https://trends.oecd-opsi.org; www.gov.uk/government/case-studies/using-data-from-electricity-meters-to-predict-energy-consumption.

Los gobiernos también están adoptando ampliamente, o planean adoptar, proyectos de IA que apoyen los servicios de asistencia social para los ciudadanos y una vida mejor para las personas (véase el Recuadro 3.11). Estos objetivos impactan en los ODS de Fin de la pobreza (ODS 1), Hambre cero (ODS 2), Salud y bienestar (ODS 3), Igualdad de género (ODS 6) y Reducción de las desigualdades (ODS 10).

Recuadro 3.11: Proyectos de IA para mejorar la calidad de vida

Decisiones de asistencia social

Dinamarca planea desarrollar algoritmos de Aprendizaje Automático de IA para ayudar a los funcionarios públicos a tomar decisiones sobre si los ciudadanos y empresas deben recibir ayuda financiera y de otro tipo (por ejemplo, apoyo a los daneses de edad avanzada, asistencia a las familias de bajos ingresos y asistencia para la vivienda), por parte del gobierno. El gobierno considera que a través de este método se pueden tomar decisiones más precisas y objetivas, libres de prejuicios humanos. Además, puede ayudar a hacer frente al desafío del envejecimiento de la población, ya que sólo se dispone de un número limitado de funcionarios públicos

Recuadro 3.11: Proyectos de IA para mejorar la calidad de vida *(Cont.)*

para procesar un número cada vez mayor de solicitudes de asistencia social. Para que esto sea posible, el gobierno se ha centrado en dos desafíos específicos:

- Cómo establecer una legislación adecuada que permita la adopción de decisiones automatizadas.
- Hacer que las máquinas puedan leer y comprender los flujos de datos y decisiones subyacentes.

Prevención de la esclavitud desde el espacio

El centro de investigaciones del Reino Unido, Rights Lab, lanzó recientemente Esclavitud desde el Espacio (*Slavery from Space*), un proyecto para poner fin a la esclavitud moderna. Utiliza algoritmos de aprendizaje automático que analizan datos satelitales de alta resolución para estimar el número de hornos para ladrillos en el "Cinturón de ladrillos" (*Brick Belt*) del sur de Asia, una zona donde la esclavitud es muy frecuente, ayudando así a calcular el grado de la esclavitud moderna en la región. Antes de este trabajo, se desconocía la cantidad total de los hornos para ladrillos y, por inferencia, la esclavitud, lo que obstaculizaba la acción de los organismos competentes. Esta innovación proporciona datos para ayudar a las ONG y a los gobiernos a luchar contra la esclavitud moderna. El equipo de Rights Lab estima que un tercio de la esclavitud puede detectarse desde el espacio mediante este método.

Proyectos de Pulso Mundial

Pulso Mundial (*Global Pulse*) es la iniciativa principal de las Naciones Unidas sobre Big Data y consiste en una red de laboratorios de innovación. Pulso Mundial está trabajando para implementar el análisis de voz a texto basado en la IA del contenido de las radios locales, para ayudar a comprender los sentimientos locales en relación con la afluencia de refugiados. Por ejemplo, al analizar los debates de la radio local, los algoritmos de Aprendizaje Automático han descubierto información valiosa que no se había recopilado previamente a través de otros mecanismos. Se han podido identificar desastres de pequeña escala y su impacto en el público, así como las áreas de vulnerabilidad para los refugiados. Pulso Global ha puesto en marcha otros proyectos que utilizan la IA para impulsar los objetivos relacionados con los ODS.

Fuente: https://govinsider.asia/innovation/exclusive-denmark-plans-to-use-ai-for-welfare-payments; https://oecd-opsi.org/innovations/slavery-from-space; https://rightsandjustice.nottingham.ac.uk; www.unglobalpulse.org/projects/pilot-studies-using-machine-learning-analyse-radio-content-uganda-2017; www.unglobalpulse.org/projects.

De ninguna manera son las únicas formas en que la IA puede impulsar los ODS y el desarrollo internacional en general. En un informe elaborado por Alianza Internacional de Innovación para el Desarrollo (*International Development Innovation Alliance,* IDIA) titulado *Inteligencia Artificial en el Desarrollo Internacional* (*Artificial Intelligence in International Development)* ofrece un análisis adicional sobre este tema (IDIA, 2019). Además, la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID, por sus siglas en inglés) publicó recientemente Reflejar el pasado, dar forma al futuro: cómo lograr que la IA funcione al servicio del

desarrollo internacional (*Reflecting the Past, Shaping the Future: Making AI Work for International Development*)[19] para ayudar a las personas con experiencia técnica limitada a abrirse paso en el panorama de la IA en los países en desarrollo.

Mantenerse actualizado respecto a los avances de la IA en el sector público

El campo de la IA evoluciona y crece rápidamente en todos los sectores, incluyendo el público, debido al lanzamiento continuo de nuevas estrategias y proyectos gubernamentales.

Si bien el objetivo de este capítulo es ofrecer una visión general actual de las estrategias nacionales y las tendencias gubernamentales en materia de IA, la situación seguirá cambiando rápidamente. Con el fin de ayudar a los funcionarios públicos y a otros lectores interesados a actualizarse de forma continua, el OPSI actualiza con regularidad su página sobre Eestrategias de IA y Componentes del Sector Público[20] en la que se desglosa información por país, y su página de Herramientas y Recursos de IA,[21] la cual ofrece un depósito categorizado de herramientas y recursos prácticos que pueden ayudar a los funcionarios del gobierno a aprender más sobre la IA y sus posibles funciones en el sector público.

Si bien en este capítulo se busca proporcionar ejemplos ilustrativos de proyectos concretos de IA, es imposible ofrecer una lista exhaustiva, ya que cada día los gobiernos examinan nuevos proyectos. El OPSI invita a los funcionarios públicos a mantenerse al día con las últimas novedades accediendo a los siguientes recursos:

- La **Plataforma de Casos de Estudio (*Case Study Platform*)** del OPSI[22] recopila y comparte cientos de innovaciones del gobierno para ayudar a difundir buenas ideas. Cualquier innovador del sector público puede presentar innovaciones a través de la plataforma. De los más de 300 casos que hoy en día se encuentran en la plataforma, aproximadamente 30 incluyen un componente de IA.[23]
- El próximo **Observatorio de Políticas en materia de IA (AIPO, por sus siglas de inglés) de la OCDE**,[24] que se lanzará a principios de 2020, tiene como objetivo ayudar a los países a fomentar, apoyar y supervisar el desarrollo responsable de una IA fiable. Incluirá una base de datos abierta y completa de iniciativas de políticas de IA, información sobre temas de políticas públicas relacionadas con la IA (por ejemplo, empleos, aptitudes, datos, salud, transporte), y métricas e instrumentos de medición de IA (por ejemplo, métodos y mediciones de la OCDE, puntos de datos vivos de los socios).
- La Unión Internacional de Telecomunicaciones (UIT) de las Naciones Unidas ha desarrollado un **Depósito Mundial de IA** de proyectos que fomentan el progreso hacia los ODS.[25]

[19] www.usaid.gov/digital-development/machine-learning/AI-ML-in-development.
[20] https://oe.cd/aistrategies. El OPSI planea integrar en el futuro este recurso con el próximo depósito de conocimientos del Observatorio de Políticas en materia de IA de la OCDE sobre estrategias nacionales de IA para que los usuarios puedan obtener periódicamente información completa y actualizada sobre éstas.
[21] https://oe.cd/airesources.
[22] https://oecd-opsi.org/innovations.
[23] https://oecd-opsi.org/case_type/opsi/?_innovation_tags=artificial-intelligence-ai.
[24] https://oecd.ai.
[25] www.itu.int/en/ITU-T/AI/Pages/ai-repository.aspx.

- El **Kit de Herramientas para el Gobierno** D**igital (*Digital Government Toolkit*)** de la OCDE ofrece recursos sobre buenas prácticas de gobierno digital por países, incluyendo una gran cantidad sobre la gestión de datos como activo.[26]

Por último, si bien a través de este capítulo se trata de demostrar que la Inteligencia Artificial *puede* ayudar a impulsar la innovación en las políticas y servicios gubernamentales, es importante señalar que la IA no es la solución para todos los problemas. Los funcionarios públicos y todos los niveles deben tener en cuenta numerosas consideraciones al evaluar el uso de la IA. El OPSI promueve la experimentación con la IA, según corresponda y de una manera informada. Además, aunque la IA puede ser una fórmula perfecta para un problema particular, su potencial depende de que el gobierno se comprometa e invierta. Durante décadas se han llevado a cabo debates similares en el gobierno sobre las innovaciones digitales, muchas veces con retrasos en los resultados o resultados poco relevantes en el sector público. En el siguiente capítulo se brinda orientación sobre la forma en que los gobiernos pueden determinar si la IA es adecuada para sus problemas y qué pueden hacer para aprovechar las oportunidades que ofrece.

[26] https://oecd.org/governance/digital-government/toolkit/goodpractices.

Capítulo 4

Reflexiones y orientación para el sector público

"A menudo digo a mis estudiantes que no se dejen engañar por el nombre de 'inteligencia artificial' - no hay nada artificial en ella… La I.A. está hecha por humanos, su comportamiento está dictado por humanos y, en última instancia, está hecha para tener un impacto en las vidas de los humanos y su sociedad." - FeiFei Li, Profesora de Informática en la Universidad de Stanford, Exjefe Científico de AI/ML y Vicepresidenta de Google Cloud.

Como se ha mostrado en capítulos anteriores, existe un potencial significativo para la aplicación de la IA en el sector público. También existe una gran cantidad de desafíos y repercusiones que los dirigentes gubernamentales y los funcionarios públicos deben tener en cuenta al momento de determinar si la IA puede ayudarles a resolver distintos problemas y cumplir sus misiones. La obtención de apoyo dependerá de que se establezcan una dirección y una narrativa claras para el uso de la IA en el sector público, para servir mejor a los ciudadanos y a las empresas. Los gobiernos también deben garantizar espacio suficiente para permitir la flexibilidad y la experimentación a fin de facilitar un aprendizaje rápido.

Cabe señalar que los gobiernos deberán desarrollar formas de determinar si la IA es la mejor solución para un problema determinado, y ofrecer conductos para identificar y prestar atención a esos problemas. Un factor crucial para lograrlo es comprender las necesidades de sus poblaciones. Como factor transversal, también tendrán que reunir equipos multidisciplinarios y diversos para ayudar en esas determinaciones y a promover el desarrollo de iniciativas y proyectos de IA que sean a la vez eficaces y éticos.

Una vez que los gobiernos han decidido aprovechar la IA, como muchos gobiernos y organismos internacionales han reconocido, es fundamental que desarrollen una estrategia fiable, justa y responsable para diseñar y aplicar una IA que identifique las concesiones necesarias, mitigue los riesgos y sesgos, y garantice una función adecuada para los humanos.

Una parte clave de esto será asegurar un enfoque constante en los usuarios y los individuos que pueden verse afectados por los sistemas de IA a lo largo de su ciclo de vida completo. Los gobiernos también deben considerar los elementos fundamentales que hacen posible la innovación impulsada por la IA. Los datos son los elementos fundamentales para la IA, y para desplegarla de manera eficaz se necesita una estrategia de datos clara que permita a los gobiernos acceder a datos robustos y precisos, de una manera que se mantenga la privacidad y que se ajuste a las normas sociales y éticas.

Los gobiernos también necesitarán tener acceso a profesionistas de primera categoría, a los productos, servicios e infraestructuras esenciales, tanto en el sector público como en el privado. Tendrán que determinar las proporciones correctas para la creación de capacidad interna, como decidir entre la creación de equipos internos de ciencia de datos, frente a la subcontratación del desarrollo de capacidades de IA al sector privado o a otros socios externos. Estas decisiones son específicas para cada país y contexto, y pueden influir en el enfoque adoptado hacia los proyectos de IA y los costos de los mismos. Independientemente de ello, es importante que los funcionarios públicos tengan, como mínimo, un nivel básico de conocimientos sobre datos y de comprensión de la ciencia de datos y las herramientas relacionadas con ésta a medida que se vuelven cada vez más comunes en la vida cotidiana y, en cierta medida, obligatorias para el futuro de la función pública. Así pues, debe prestarse especial atención a ofrecer oportunidades a los actuales funcionarios públicos para que desarrollen esas capacidades, así como a considerar las competencias que cabe esperar de los futuros funcionarios públicos.

Por último, aunque los problemas apremiantes de la actualidad suelen ser prioritarios, los gobiernos también deben reconocer los cambios potencialmente importantes que la IA podría traer en el futuro. Deben buscar métodos para explorar y anticipar las posibilidades del mañana para poder actuar hoy.

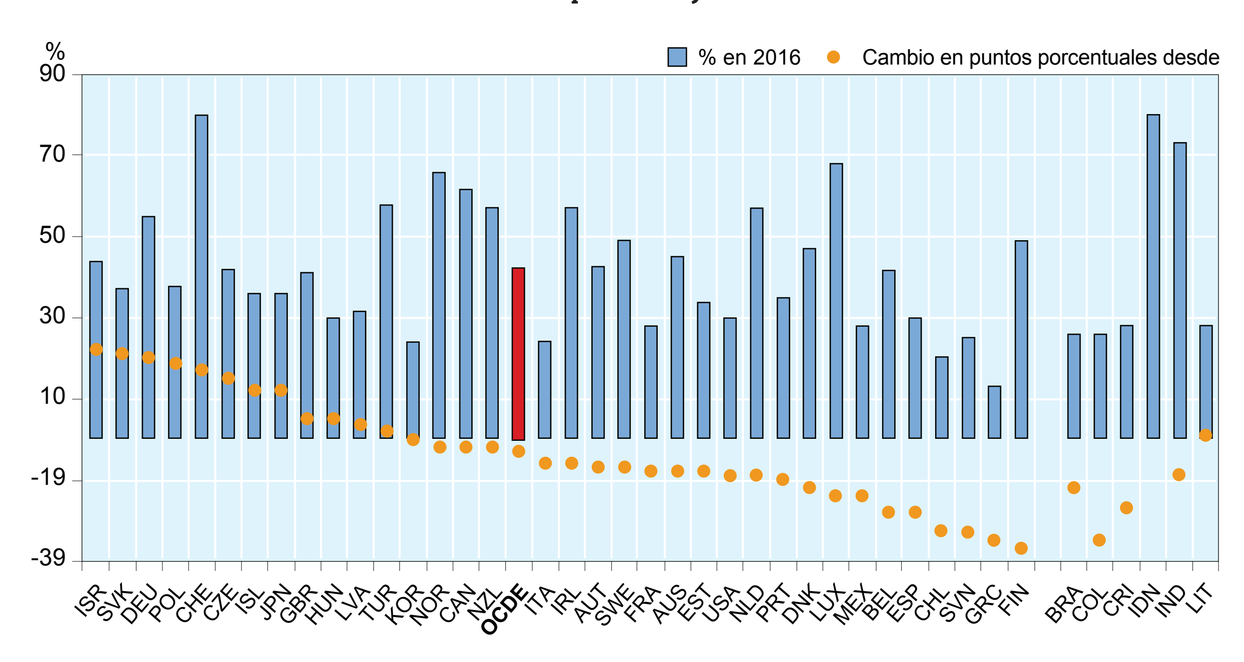
En esta sección se exploran estas cuestiones con el fin de ayudar a los dirigentes gubernamentales y a los funcionarios públicos a maximizar los beneficios de la IA, aprender de las acciones de los demás y minimizar los posibles riesgos. Concluye estableciendo un marco de trabajo para ayudarles a reflexionar sobre su acercamiento al uso de la IA para la innovación del sector público.

Proporcionar apoyo y una dirección clara, así como espacio para permitir la flexibilidad y la experimentación

Ante la continua demanda pública y las presiones relativas a los recursos, la Inteligencia Artificial presenta una importante oportunidad para mejorar la productividad y la calidad de los servicios públicos y las operaciones gubernamentales. No obstante, los bajos niveles de confianza en el gobierno (véase la Figura 4.1) ponen énfasis en la necesidad de que el sector público marque la pauta adecuada desde los niveles más altos y adopte un enfoque que haga hincapié en una IA fiable, ética y equitativa. Investigaciones recientes del Boston Consulting Group indican específicamente que el apoyo a la IA gubernamental se correlaciona con la confianza en el gobierno, y que “la confianza en las instituciones es esencial para que los gobiernos obtengan el apoyo necesario para desplegar las capacidades de IA” (Carrasco et al., 2019). Esta pauta también puede servir como factor facilitador, ya que quienes ocupan los altos mandos tienen el poder de establecer una dirección estratégica que puede repercutir en cada nivel inferior, ayudando de esta forma a establecer la cultura

en general (OCDE, 2017b). Sin embargo, no es suficiente con sólo dejar esto en la teoría. Estos líderes también están en posición de crear espacio para permitir la flexibilidad y la experimentación, como laboratorios y espacios aislados seguros.

Figura 4.1: La confianza en el gobierno ha ido en declive, a menudo desde un punto de partida bajo



Fuente: World Poll de Gallup

Marcar la pauta

Se requerirá un apoyo político sostenido y de alto nivel para crear un entorno estable y propicio que permita que maduren las estrategias y soluciones de la IA. La pauta marcada por los más altos niveles de gobierno tiene una función de convocatoria crucial para establecer la dirección del desarrollo de la IA y su uso en la sociedad en general. Esta pauta también envía señales a los funcionarios públicos en todos los niveles, además de brindarles una cobertura de primer nivel, lo cual les permite impulsar la innovación y el progreso. Como vimos en el Capítulo 1, la creciente capacidad de soluciones basadas en enfoques participativos, posibilitada gracias a la democratización de la potencia de procesamiento y las plataformas de colaboración, entre otros, contribuye a la actual oleada de interés por la IA. Habilitar el potencial de estas soluciones es igual de relevante en el sector público, y la cobertura de primer nivel del liderazgo puede ayudar en este sentido.

Hay muchas trayectorias posibles para la IA. Los gobiernos deben asegurarse de que se utilice de manera que promueva y proteja los objetivos y valores de la sociedad (Mateos-Garcia, 2018). Los planes de despliegue dentro del gobierno también deben ser compatibles con el impulso de la innovación a través de la I+D, además de respaldar los planes para dicha innovación, así como con el fomento de la IA en la economía en general a través de la inversión en infraestructura y habilidades, el entorno normativo más amplio y otras políticas de estrategia industrial. Si bien uno de los objetivos principales de los gobiernos

puede ser utilizar la IA para mejorar las políticas y los servicios públicos, también deberían considerar su papel en la adopción de un enfoque de "estado emprendedor" para impulsar el crecimiento y la innovación, utilizando todos los instrumentos a su disposición para configurar los mercados y asumir riesgos para lograr su visión (Mazzucato, 2013). El Gobierno de los Estados Unidos, por ejemplo, ofrece este tipo de apoyo de alto nivel a través de su visión para mantener el liderazgo en materia de IA, como se expone en el Recuadro 4.1.[1]

Recuadro 4.1: Decreto Ejecutivo del Presidente de los Estados Unidos sobre el mantenimiento del liderazgo nacional en materia de Inteligencia Artificial [*The President of the United States Executive Order on Maintaining American Leadership in Artificial Intelligence*]

El Decreto Ejecutivo del Presidente ofrece un ejemplo de una visión clara de la IA y de cómo beneficiará el crecimiento económico de los Estados Unidos, los intereses de seguridad y la vida de sus ciudadanos. Expresa los objetivos de mantener el liderazgo mundial de los Estados Unidos y de asegurar que la IA evolucione "de manera compatible con los valores, políticas y prioridades de nuestra nación". Luego explica cómo se utilizarán los mecanismos del Gobierno Federal para lograr estos objetivos a través de:

- la promoción y el financiamiento de la I+D para impulsar avances tecnológicos
- el desarrollo de normas técnicas pertinentes y el fomento de la experimentación para aumentar el despliegue de la IA
- la creación de competencias para desarrollar y aplicar tecnologías de IA, incluso entre la fuerza de trabajo federal
- el fomento de la confianza en la IA garantizando que proteja la privacidad y las libertades individuales
- la creación de un entorno internacional que cree mercados para las empresas de inteligencia artificial de EE. UU. y que proteja la seguridad y los intereses estratégicos de EE. UU.

Fuente: www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence.

Además de contar con un apoyo político de alto nivel, los gobiernos tendrán que articular una visión convincente sobre las formas en las que la IA puede transformar los servicios y operaciones públicos para beneficiar a los ciudadanos y a las empresas, al mismo tiempo que se mantiene la confianza del público. En el Capítulo 3 se señala que la mayoría de los países con estrategias nacionales de IA incluyen un método que involucra específicamente a la IA para la innovación y la transformación del sector público, y algunos cuantos incluso tienen una estrategia explícitamente dedicada al gobierno. Por ejemplo, la Estrategia AuroraAI de Finlandia (véase el caso de estudio que se incluye en el Anexo A) expresa claramente el

[1] Es posible encontrar detalles adicionales sobre la estrategia de los EE. UU. y las estrategias para la IA de otros países en un suplemento digital de este informe en https://oecd-opsi.org/ai.

ambicioso objetivo de desarrollar una sociedad centrada en los humanos y basada en el bienestar integral de su población, sus empresas y la sociedad como un todo. Sin embargo, la mayoría de los países carecen de una estrategia de IA o de un enfoque centrado en el sector público. El desarrollo de estrategias y el dar prioridad a los casos de uso práctico que demuestren las formas en las que la IA puede mejorar los servicios para los ciudadanos, pueden sentar la base para obtener el apoyo por parte del público.

Cada estrategia y enfoque nacional debe operar dentro de su propio contexto único y dentro de su propia cultura y normas. Los gobiernos deberían entablar un diálogo deliberativo con los ciudadanos y las empresas que les permita comprender más claramente sus perspectivas y valores (Balaram, Greenham y Leonard, 2018). En particular, los usuarios de los servicios públicos podrían desear una participación significativa y garantías sobre la forma en que el uso de la IA repercutirá en los servicios de los que dependen. En algunos casos, ellos también podrían convertirse en cocreadores de servicios públicos con IA, lo que por definición implica una participación significativa de los usuarios (Lember, Brandsen y Tõnurist, 2019). Por último, la IA tiene el potencial de ayudar a los gobiernos a avanzar hacia la existencia de servicios públicos proactivos. Esos servicios anticipan y atienden las necesidades del usuario antes de que éste tenga que intervenir (p. ej., llenar un formulario) (Scholta et al., 2019). Esto no puede lograrse si no se cuenta con una comprensión muy profunda de las necesidades de los usuarios, cuya clave es contar con su participación. Esta participación será cada vez más importante y debería incluirse como parte integral de las estrategias nacionales y la dirección general. Los funcionarios públicos también deben estar facultados para interactuar con los usuarios.

Del mismo modo, para asegurar y mantener el apoyo dentro del sector público será necesario contar con un discurso claro que explique cómo la IA puede ayudar a los empleados del sector público a prestar mejor los servicios, reducir la cantidad de tiempo que dedican a las tareas rutinarias y permitirles centrarse en tareas de mayor valor en las que puedan tener un impacto más profundo.[2] Si no se mitiga, la resistencia de los trabajadores del sector público podría frenar el despliegue de la IA, limitar su eficacia y dañar la moral. Es poco probable que presentar argumentos a favor de la IA basados en su potencial para reducir el número de empleados obtenga el respaldo deseado, además de que no es fidedigno, ya que es poco probable que la IA reemplace a los trabajadores del sector público en el futuro inmediato. La Academia Digital Canadiense (véase el Recuadro 4.22) ofrece un ejemplo de un enfoque innovador para mejorar los conocimientos de los funcionarios públicos sobre la IA.

Es importante señalar que no es necesario que cada uno de los gobiernos se encargue de todos los aspectos del desarrollo de programas y ecosistemas sólidos para la IA. En cambio, pueden aprovechar las oportunidades para colaborar internacionalmente en el diseño de estrategias y normas de la IA (Mateos-Garcia, 2018). Muchos gobiernos se enfrentan a los mismos problemas en torno a la IA, y existen grandes oportunidades de trabajar en conjunto para abordarlas y explorar normas comunes y enfoques de colaboración. Los Principios de la OCDE sobre IA (Recuadro 4.2) ofrecen el primer conjunto de normas internacionales acordadas por los gobiernos. De manera similar, las "Directrices Eticas para una Inteligencia Artificial fiable" (véase el caso de estudio del Anexo A) formulan una serie de principios para fomentar y garantizar una IA robusta y ética (European Commission [*Comisión Europea*], 2019a).

2 www.digital.nsw.gov.au/digital-transformation/policy-lab/artificial-intelligence.

Recuadro 4.2: Los Principios de la OCDE sobre Inteligencia Artificial

Los Principios de la OCDE sobre Inteligencia Artificial respaldan una IA que sea innovadora y fiable, y que respete los derechos humanos y los valores democráticos. Los países miembros de la OCDE adoptaron los principios el 22 de mayo de 2019 como parte de la Recomendación del Consejo sobre Inteligencia Artificial de la OCDE. Estos principios establecen normas para la IA que son lo suficientemente prácticas y flexibles para resistir el paso del tiempo en un campo que evoluciona rápidamente. Complementan las normas vigentes de la OCDE en ámbitos como la privacidad, la gestión de riesgos de seguridad digital y la conducta empresarial responsable.

La Recomendación identifica cinco principios complementarios basados en valores para la gestión responsable de una IA fiable:

- La IA debería beneficiar a las personas y al planeta a través del impulso del crecimiento inclusivo, el desarrollo sostenible y el bienestar.
- Los sistemas de IA deberían diseñarse de manera que respeten el estado de derecho, los derechos humanos, los valores democráticos y la diversidad, y deberían incluir las salvaguardas adecuadas, por ejemplo, permitir la intervención humana cuando resulte necesario, para garantizar una sociedad justa y equitativa.
- Es necesario que exista transparencia y una divulgación responsable en torno a los sistemas de IA para asegurar que las personas entiendan los resultados obtenidos con la IA y puedan cuestionarlos.
- Los sistemas de IA deben funcionar de manera robusta, segura y protegida a lo largo de su ciclo de vida, y se debe evaluar y gestionar los posibles riesgos de forma continua.
- Las organizaciones y las personas que desarrollen, desplieguen u operen sistemas de IA deben ser responsables de su correcto funcionamiento en concordancia con los principios antes señalados.

La Recomendación también formula cinco sugerencias a los encargados de dictar políticas nacionales y en materia de cooperación internacional para una IA fiable.

Éstas se analizan en el Recuadro 1.10, que se encuentra en secciones anteriores del documento.

Fuente: http://oecd.ai.

Crear espacio para la experimentación

La experimentación y el aprendizaje iterativo son cruciales para desarrollar las capacidades de la IA. Si los profesionales no tienen la libertad de probar nuevas formas de desarrollar y prestar servicios, es poco probable que se alcance el potencial de la IA en los servicios y operaciones públicos. Sin embargo, la adopción de un enfoque experimental en torno al uso de la IA podría contrarrestar los esfuerzos por establecer sistemas robustos y procesos compatibles en todo el gobierno. Por otra parte, es probable que el despliegue de los sistemas

de IA sea lento si los encargados de la toma de decisiones esperan hasta que se cuente con marcos y normas de gobernanza ideales. En resumen, los gobiernos deben dedicar tiempo y espacio a la experimentación, como ha hecho Nueva Zelanda (véase el Recuadro 4.3); de otra forma, es posible que no se le dé prioridad a la IA por encima de las urgentes presiones cotidianas. Sin experimentación y aprendizaje abiertos, existe el riesgo de que se arraiguen y normalicen prácticas poco éticas o descuidadas, lo que podría llevar a trayectorias a largo plazo subóptimas o incluso peligrosas.[3]

Recuadro 4.3: El Laboratorio de Innovación de Servicios [*Service Innovation Lab*] y la Iniciativa Mejores Normas [*Better Rules*] de Nueva Zelanda

El Laboratorio de Innovación de Servicios es un espacio neutral para todo el gobierno que permite a los organismos del sector público colaborar en innovaciones que faciliten el acceso del público a los servicios gubernamentales. Aunque no se centra estrictamente en la IA, opera como un laboratorio de diseño y desarrollo que permite experimentar, impulsar y facilitar un cambio sistémico en el gobierno en beneficio de la sociedad, centrado en las necesidades de los usuarios. La labor del laboratorio también se enfoca en dirigir el financiamiento público hacia mejoras sistémicas, esfuerzos horizontales en torno a objetivos compartidos, componentes reutilizables de alto valor e innovación viable para todos los organismos del sector público participantes.

El Laboratorio de Innovación de Servicios colabora con dependencias y socios de toda Nueva Zelanda para promover una mayor innovación en toda la administración pública. Si bien no se centra en la IA, el Laboratorio es un ejemplo de trabajo colaborativo interinstitucional para experimentar, superar las barreras sistémicas que impiden la innovación y crear prototipos de nuevos enfoques para la prestación de servicios integrados que se diseñan en torno a las necesidades de los usuarios. Por lo tanto, ofrece un ejemplo de cómo los gobiernos pueden adoptar un enfoque ágil y adaptable a la innovación sistémica.

Por ejemplo, el proyecto Mejores Normas del Laboratorio reescribe las leyes en forma de código consumible por máquina para ayudar a garantizar la implementación correcta, y a desarrollar bucles de retroalimentación en tiempo real entre el diseño legislativo y el proceso de implementación. Al funcionar como la única fuente de datos legibles por máquina, dicho código puede servir de base para modelos y algoritmos de IA. Si las leyes cambian, dichos cambios pueden reflejarse de inmediato y de forma precisa en el algoritmo para ayudar a garantizar la correcta implementación.

Fuente: https://oecd-opsi.org/innovations/the-service-innovation-lab, https://trends.oecd-opsi.org.

En algunas ocasiones, podría ser necesario colocar los esfuerzos de la IA dentro de áreas cerradas para la experimentación. El desarrollo del Modelo de Facetas de la Innovación [*Innovation Fascets*] del OPSI[4] ha puesto de manifiesto una serie de características dinámicas de las diferentes perspectivas sobre la innovación. Por ejemplo, gran parte del trabajo de los gobiernos en cuanto a la exploración de la IA entra en la categoría de innovación

[3] www.slideshare.net/JuanMateosGarcia/d4p-complex-economicsaiv2.
[4] https://oecd-opsi.org/projects/innovation-facets.

anticipada.[5] Esta faceta implica explorar e involucrarse en problemas emergentes que podrían determinar las prioridades y compromisos futuros. También tiene el potencial de subvertir los paradigmas existentes. Por lo general, las ideas muy nuevas no cohabitan bien con las estructuras, procesos, flujos de trabajo y normas de presentación de informes existentes, ya que aún es necesario concretar los detalles específicos de cómo funcionará la idea en la práctica. Por lo tanto, en general la innovación anticipada debe estar protegida de las actividades principales y contar con su propia autonomía. De lo contrario, es probable que las presiones de las prioridades tangibles existentes terminen consumiendo los recursos necesarios, o que el concepto choque con normas que no hayan tenido en cuenta sus posibilidades.

Varios gobiernos están tratando de lograr esto mediante la creación de “espacios aislados”. Este enfoque les permite llevar a cabo experimentos en espacios seguros y apartados que ayudan a fomentar la innovación, a la vez que aprenden sobre nuevos modelos y las formas de manejarlos. Los espacios aislados regulatorios, por ejemplo, pueden flexibilizar ciertas normas o regulaciones en función de una serie de condiciones (p. ej., limitación del tiempo y del número de participantes) (Eggers, Turley y Kishnani, 2018). Estos espacios aislados también pueden ayudar a mejorar la seguridad y la privacidad de los datos, ya que representan un espacio seguro supervisado en el que los datos pueden separarse de otras funciones y redes (CIPL, 2019). En estos espacios seguros, los funcionarios pueden aprender más sobre los datos, el potencial de la IA, los tipos de áreas sensibles involucradas y los métodos necesarios para protegerlos y garantizar la protección de la privacidad de las personas. Si bien suelen estar orientados al sector privado, cada vez se considera más la posibilidad de utilizar los espacios aislados para la IA del sector público (véase el Recuadro 4.4).

Recuadro 4.4: Espacios aislados para la Inteligencia Artificial en el sector público

Estonia

En julio de 2019, Estonia adoptó la Estrategia Nacional de Inteligencia Artificial de Estonia (estrategia “Kratts”). Como parte de esta estrategia, los planes gubernamentales para el desarrollo de proyectos piloto se beneficiarían de un financiamiento público más flexible, y de la creación de espacios aislados para probar y desarrollar soluciones de IA dirigidas al sector público y acelerar su adopción. Estos espacios aislados pueden proporcionar flexibilidad normativa y acceso temporal a los recursos de la infraestructura de pruebas (p. ej., el procesamiento de datos de alto rendimiento). La estrategia destaca que no sólo se necesitan espacios aislados regulatorios, sino también tecnológicos.

Lituania

En abril de 2019, el Gobierno de Lituania publicó una estrategia nacional en materia de IA destinada a modernizar y ampliar el ecosistema de IA actual en Lituania, y a garantizar que este país esté preparado para un futuro con IA. Una de las recomendaciones clave es desarrollar un espacio aislado regulatorio que permita el uso y la realización de pruebas de sistemas de IA en el sector público durante un tiempo limitado. Esto permitirá a los desarrolladores probar los productos en un

[5] https://oecd-opsi.org/innovation-facets-part-6-anticipatory-innovation.

Recuadro 4.4: Espacios aislados para la Inteligencia Artificial en el sector público *(Cont.)*

entorno real, y al sector público determinar qué soluciones pueden integrarse de forma más plena.

Finlandia

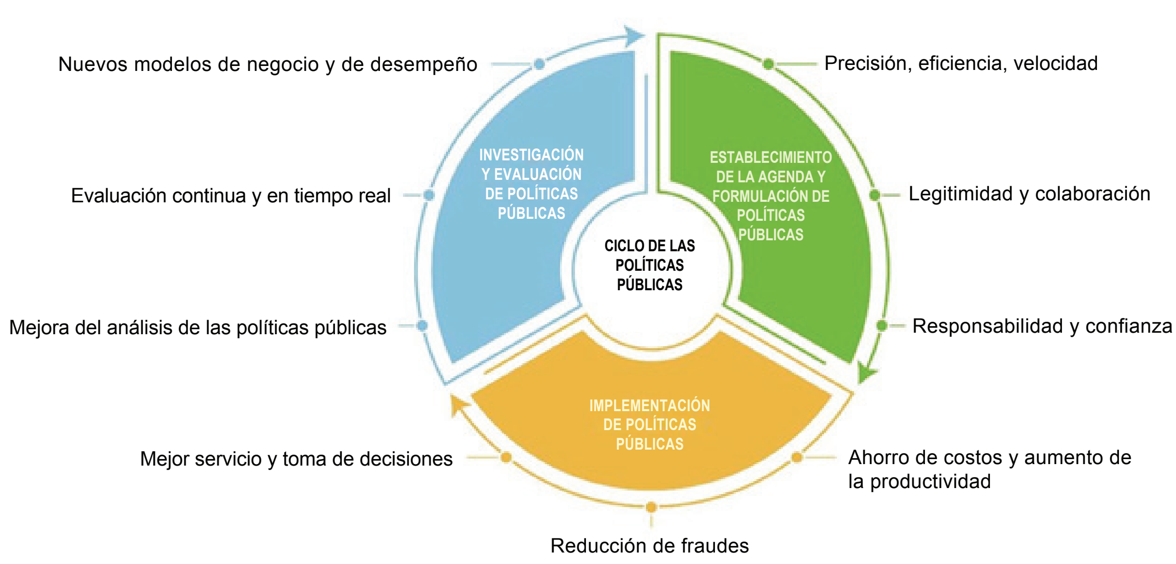
La estrategia AuroraAI de Finlandia requiere un espacio aislado regulatorio para experimentar, de forma controlada, con datos que hayan sido autorizados por los ciudadanos, así como para explorar la necesidad de realizar algún cambio legislativo para alcanzar el máximo potencial. Es posible encontrar más detalles en el caso de estudios sobre la Estrategia Nacional de Inteligencia Artificial de Finlandia que se encuentra en el Anexo A.

Fuente: https://e-estonia.com/estonia-accelerates-artificial-intelligence; www.riigikantselei.ee/sites/default/files/riigikantselei/strateegiaburoo/eesti_tehisintellekti_kasutuselevotu_eksperdiruhma_aruanne.pdf; http://kurklt.lt/wp-content/uploads/2018/09/StrategyIndesignpdf.pdf; https://vm.fi/documents/10623/1464506/AuroraAI+development+and+implementation+plan+2019%E2%80%932023.pdf.

¿Es la IA la mejor solución al problema?

La IA puede identificar patrones o irregularidades en los datos para mejorar la precisión de la toma de decisiones, asignar mejor los recursos, anticipar las necesidades no satisfechas, detectar fraudes o riesgos de seguridad, entre muchas otras cosas. Cuando se diseñan y aplican de forma adecuada, estas capacidades permiten a la IA contribuir de forma positiva a las actividades gubernamentales a lo largo de todo el ciclo de elaboración de políticas, desde el establecimiento de la agenda y la formulación de políticas públicas, hasta la implementación y la evaluación (véase la Figura 4.2).

Figura 4.2: Beneficios de la IA en cada etapa del ciclo de las políticas públicas



Fuente: Pencheva, I., M. Esteve and S.J. Mikhaylov (2018), *Big Data and AI – A Transformational Shift for Government: So, What Next for Research?* https://journals.sagepub.com/doi/pdf/10.1177/0952076718780537.

Sin embargo, la IA no siempre es la mejor solución y en muchos casos no es viable utilizarla. Un problema común con las tecnologías emergentes, como las que surgen del campo de la IA, es el riesgo de que la gente empiece a crear soluciones y *después* busque problemas para que la tecnología los resuelva. En general, los gobiernos deben intentar comprender y centrarse en los resultados que tanto ellos como los ciudadanos desean obtener y en los problemas que se interponen en el camino. Habiendo adquirido este conocimiento y estas prioridades, pueden entonces identificar si la IA (u otra herramienta distinta) es la mejor solución para ayudar a alcanzar estos objetivos (Mulgan, 2019). Por consiguiente, los gobiernos deben evaluar una serie de cuestiones para determinar si la IA es una buena opción.

Comprender las necesidades de la gente

Los gobiernos deben poseer una sólida comprensión de las necesidades de los ciudadanos, residentes, empresas, funcionarios públicos y todas aquellas personas que puedan interactuar con una solución basada en la Inteligencia Artificial o verse afectadas por ella. A menos que interactúen con los usuarios potenciales (tanto dentro como fuera del gobierno, según corresponda), los funcionarios públicos no podrán determinar con precisión cuáles son los problemas existentes y si una posible aplicación o alternativa de la IA satisfará las necesidades básicas. El Manual de Servicios Digitales de Estados Unidos [*United States Digital Services Playbook*] ofrece una lista de verificación y un conjunto de preguntas clave para descubrir estas necesidades (Recuadro 4.5). El Manual de Habilitación de TIC [*ICT Commissioning Playbook*], elaborado por el Grupo Temático de Líderes de Gobierno Electrónico (E-Leaders) de la OCDE sobre la habilitación de las TIC, también se centra en la comprensión de las necesidades de los usuarios, entre otros temas.[6]

Recuadro 4.5: Manual de Servicios Digitales de Estados Unidos – Entender lo que la gente necesita

Lista de verificación

- Durante las primeras etapas del proyecto, pasar tiempo con los usuarios actuales y posibles futuros usuarios del servicio.
- Utilizar una variedad de métodos de investigación cualitativos y cuantitativos para determinar los objetivos, las necesidades y los comportamientos de las personas; ser prudente respecto al tiempo empleado.
- Someter a prueba los prototipos de soluciones con gente real; llevar a cabo como estudio de campo si es posible.
- Documentar los hallazgos relativos a los objetivos, necesidades, comportamientos y preferencias de los usuarios.
- Compartir los hallazgos con el equipo y la dirección de la dependencia gubernamental.
- Crear una jerarquía con las tareas que el usuario está tratando de realizar, también conocidas como “historias de usuario”.

[6] https://playbook-ict-procurement.herokuapp.com.

Recuadro 4.5: Manual de Servicios Digitales de Estados Unidos – Entender lo que la gente necesita *(Cont.)*

- A medida que el servicio digital se va construyendo, hacer pruebas con los posibles futuros usuarios con regularidad para garantizar que satisfaga las necesidades de las personas.

Preguntas clave

- ¿Quiénes son sus usuarios principales?
- ¿Qué necesidades de los usuarios atenderá este servicio?
- ¿Por qué el usuario quiere o necesita este servicio?
- ¿Quiénes tendrán más dificultades con el servicio?
- ¿Qué métodos de investigación se utilizaron?
- ¿Cuáles fueron los hallazgos clave?
- ¿Cómo se documentaron los hallazgos? ¿Dónde pueden acceder los futuros miembros del equipo a la documentación?
- ¿Con qué frecuencia se realizan pruebas con gente real?

Fuente: https://playbook.cio.gov.

Para muchos de los desafíos digitales del sector público, las soluciones más pertinentes suelen ser los usos sencillos pero eficaces de las tecnologías existentes y la mejora de la interoperabilidad, incluso con los sistemas heredados. Por ejemplo, la empresa de reciente creación del Reino Unido, Accurx, se propuso originalmente utilizar Aprendizaje Automático para mejorar la eficacia de los antibióticos recetados (p. ej., para ayudar a prevenir la resistencia a los antibióticos), pero descubrió que un modelo comercial más eficaz implicaba el uso de mensajes de texto para aumentar el número de pacientes que acudían a las citas médicas (Lewin, 2019). Si no se conoce a los usuarios, sus necesidades y cómo las soluciones tecnológicas pueden encajar en sus vidas, obtener estos tipos de conocimientos se vuelve una tarea más difícil.

Considerar la capacidad de la IA para resolver problemas y obtener resultados

Una vez que se está consciente de las necesidades de los usuarios, la siguiente etapa para evaluar eficazmente la idoneidad de la IA es el diagnóstico y la definición del problema. Este proceso comienza generalmente con el desglose de las actividades o servicios pertinentes hasta llegar a sus tareas fundamentales, e identificando si la IA puede realizarlas o prestarlas como servicio de forma más eficaz. Es en ese momento que la automatización de las tareas puede volverse una prioridad en función de aquellas que tendrán mayor impacto en la rentabilidad del servicio. Es probable que una IA bien diseñada realice mejores predicciones que los humanos en los casos en los que el factor de las interacciones complejas entre muchos indicadores mejore la predicción; cuando haya un gran volumen de datos estables y representativos (que permitan que las interacciones sean un buen predictor de eventos futuros); y cuando las predicciones sean rutinarias y no muy poco frecuentes (Agarwal, Gans y Goldfarb, 2018).[7]

[7] https://faculty.ai/products-services/ai-strategy.

Sin embargo, la IA podría no ser la solución óptima para muchos, o incluso para la mayoría de los problemas. Es necesario realizar un análisis cuidadoso de las capacidades de las herramientas específicas de IA para determinar si deben ser una parte o toda la solución de un desafío específico. En el Capítulo 2 se exponen las capacidades de diversas herramientas de IA y se explica qué clases de problemas podrían ayudar a resolver.

Un enfoque riguroso en el uso de la IA únicamente cuando es probable que proporcione la solución óptima a un problema específico, reducirá el riesgo de una adopción inadecuada en áreas en las que no añadirá valor. El Gobierno del Reino Unido ha elaborado guías para evaluar si la IA es la solución correcta (véase el Recuadro 4.6).

Recuadro 4.6: Guía del Gobierno del Reino Unido para evaluar si la IA es la solución correcta

El Gobierno del Reino Unido ha creado una guía para los funcionarios con el fin de ayudarles a determinar si la IA les ayudará a cubrir las necesidades de los usuarios. Recomiendan que los funcionarios consideren si:

- los datos disponibles contienen la información requerida
- es ético y seguro utilizar los datos, y si es compatible con el Marco Ético de Datos [*Data Ethics Framework*] del Gobierno
- hay una cantidad suficiente de datos para que la IA aprenda
- la tarea es demasiado grande y repetitiva para que un humano la emprenda sin dificultad
- la IA proporcionará información que un equipo podría utilizar para obtener resultados en el mundo real.

La guía también recomienda evaluar el nivel actual de competencias y los datos existentes, seleccionar la herramienta de IA más adecuada para resolver el problema en cuestión, y luego decidir si la solución se crea o se compra.

Fuente: www.gov.uk/guidance/assessing-if-artificial-intelligence-is-the-right-solution; www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework.

Las estrategias de IA exitosas requieren mecanismos diseñados para plantear o identificar problemas específicos que la IA tiene el potencial de resolver. Los gobiernos pueden adoptar una serie de enfoques diferentes para adaptar los recursos a los problemas. A continuación, se presentan dos enfoques opuestos:

- **Enfoques descentralizados y orientados a la demanda.** Los gerentes empresariales o el personal de línea que desempeñe funciones operativas identifican los problemas que la IA puede ayudar a resolver, y recurren a los expertos internos para impulsar la transformación del servicio. Este enfoque podría facilitar la adaptación iterativa orientada a la solución de problemas, pero no conduciría necesariamente a una priorización efectiva o a enfoques uniformes en todos los gobiernos (Andrews, 2018).

- **Liderazgo transformacional centralizado.** Se trazan mapas de las posibles aplicaciones de la IA en todo el gobierno, y la experiencia y la atención se orientan hacia aquellas áreas y problemas que tienen una mayor probabilidad de beneficiarse de la IA. Esto permitiría la uniformidad, la priorización y los enfoques sistémicos, pero podría causar que los administradores de servicios adoptaran la IA como solución en lugar de centrarse en los problemas y las oportunidades perdidas que se perciben mejor desde abajo.

Las soluciones intermedias que contrarrestan algunas de las debilidades de estas dos opciones incluyen:

- Misiones o desafíos determinados centralmente para los que los expertos dentro y fuera del gobierno puedan proponer soluciones.
- La promoción y asignación de recursos a comunidades de interés o a redes de profesionales, permitiéndoles de esta manera colaborar y compartir conocimientos especializados más allá de los límites de la organización.
- La creación de fondos centrales o equipos de expertos en IA, para luego exhortar a los administradores de servicios a que identifiquen las áreas fructíferas para la exploración de la IA y a hacer una oferta por su tiempo o recursos.

En el Recuadro 4.7 se presentan ejemplos que ilustran estos enfoques.

Recuadro 4.7: Estrategias gubernamentales que vinculan los principales desafíos con soluciones tecnológicas

Misiones y grandes desafíos

El Gobierno del Reino Unido creó un reto GovTech para empresas que diseñan servicios tecnológicos para el gobierno de 20 millones de libras esterlinas, para incentivarlas a ofrecer soluciones innovadoras a los problemas del sector público. Estas estrategias orientadas a la misión alientan a las pequeñas empresas tecnológicas emergentes a crear y desarrollar soluciones innovadoras para los servicios públicos. Una vez comprobada su eficacia, las soluciones pueden escalarse para adaptarlas al mercado y a la sociedad.

En la primera ronda de GovTech Catalyst se otorgó financiamiento a través de cinco concursos diferentes: automatización de la identificación y catalogación de la propaganda de Daesh en imágenes fijas en línea, seguimiento de residuos a través de la cadena de residuos, lucha contra la soledad y el aislamiento rural, reducción de la congestión del tráfico y despliegue de sensores inteligentes en los vehículos del ayuntamiento para mejorar los servicios. De estos concursos, cinco empresas provenientes de todo el Reino Unido recibieron hasta GBP 80,000, para desarrollar soluciones digitales innovadoras para hacer frente al reto de rastrear los residuos desde su origen hasta el tratamiento y la eliminación final. El enfoque tecnológico de este financiamiento abarca más tecnologías que la Inteligencia Artificial, pero una de las cinco empresas exitosas emplea IA en su solución.

El Gobierno de Francia también ha seguido un modelo enfocado hacia los grandes desafíos como parte de los esfuerzos por implementar su estrategia nacional de IA.

Recuadro 4.7: Estrategias gubernamentales que vinculan los principales desafíos con soluciones tecnológicas *(Cont.)*

La iniciativa "Desafíos para la IA" [*Challenges AI*] promueve la innovación abierta entre organismos del sector público que se enfrentan a retos digitales, y empresas de nueva creación y pequeñas y medianas empresas (PYME) con el fin de desarrollar tecnologías innovadoras.

Comunidades de interés y redes

La Oficina de Inteligencia Artificial [*Office for AI*] y el Servicio Digital Gubernamental [*Government Digital Service*] del Gobierno del Reino Unido han elaborado directrices para los administradores de servicios sobre cómo evaluar, planificar y gestionar la IA en los servicios públicos y la administración.

La Oficina de Tecnología Ciudadana Emergente de los Estados Unidos [*Emerging Citizen Technology Office*] (ECTO, por sus siglas en inglés) trabaja con funcionarios públicos de diversas dependencias gubernamentales, así como con empresas y organizaciones cívicas, para desarrollar iniciativas de modernización de los servicios públicos en todo el gobierno. En ellas se evalúan los posibles casos de uso y se trabaja con socios para crear "recursos compartidos para la posible adopción de la tecnología".

Equipos o fondos centrales con propuestas de enfoques participativos

El Equipo de Expertos en IA del Gobierno de Estonia se dio a la tarea de analizar su marco jurídico vigente para determinar si ofrece suficiente claridad y protección en el contexto de la IA, y elaboró un plan de acción para promover el uso de la IA en todo el gobierno.

En Portugal, el gobierno lanzó una Iniciativa Nacional de Competencia Digital, "Portugal INCoDe.2020", que invertirá 10 millones de euros en los próximos tres años. El objetivo del financiamiento es estimular el uso de la ciencia de datos y la IA en el sector público. Los equipos interesados en el gobierno pueden solicitar financiamiento a través de procesos de convocatorias de licitación abiertos y competitivos. Algunos de los primeros proyectos que recibieron financiamiento, pretenden desarrollar modelos basados en la IA para predecir el riesgo de desempleo a largo plazo y detectar patrones anormales en la receta de antibióticos. Hasta agosto de 2019 se habían presentado y aprobado 44 proyectos en el marco del programa.

El Fondo de Modernización Tecnológica del Gobierno de los Estados Unidos [*Technological Modernization Fund*] (TMF, por sus siglas en inglés) es un nuevo modelo para el financiamiento de proyectos de modernización tecnológica. Las dependencias gubernamentales pueden presentar propuestas de financiamiento y conocimientos técnicos a una Junta del TMF integrada por altos dirigentes gubernamentales en materia de TI. Las propuestas se evalúan en función de:

- su impacto en la misión de la dependencia (mejora de los resultados para los usuarios y la seguridad)
- la viabilidad (incluida la capacidad de la dependencia)

Recuadro 4.7: Estrategias gubernamentales que vinculan los principales desafíos con soluciones tecnológicas *(Cont.)*

- generación de oportunidades (posibles ahorros de costos y mejoras en la calidad del servicio)
- soluciones comunes (sustitución de sistemas inseguros y anticuados por plataformas escalables que podrían ser utilizadas por otras dependencias).

El Fondo posibilita al gobierno centrar sus esfuerzos en las áreas en las que puede obtener el máximo beneficio público, al dar prioridad a las soluciones tecnológicas para mejorar la prestación de servicios y proyectos cruciales para las misiones que puedan servir como soluciones comunes y/o inspirar su reúso. Aunque su ámbito de actuación es más amplio que la IA, los funcionarios estadounidenses han animado a las dependencias a presentar propuestas para proyectos de modernización impulsados por tecnología emergente.

Fuente: www.gov.uk/government/news/smart-tracking-of-waste-across-the-uk-govtech-catalyst-competition-winners-announced, www.gov.uk/government/collections/govtech-catalyst-information, www.entreprises.gouv.fr/numerique/france-terre-d-intelligence-artificielle, https://emerging.digital.gov/TMF, https://tmf.cio.gov,https://investinestonia.com/artificial-intelligence, www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector, https://emerging.digital.gov/what-we-do, officials from the Government of Portugal.

Cuando se identifica un problema, los gobiernos pueden considerar distintas aproximaciones, entre las que se puede incluir la IA y otras soluciones. Si se identifica una solución de IA viable, los gobiernos deben evaluar si es el medio óptimo para lograr los objetivos de las políticas públicas y generar valor público, Wingfield et al. (2016). El Gobierno de Nueva Gales del Sur, en Australia, propone tres preguntas que las dependencias gubernamentales deberían plantearse al considerar la adopción de tecnologías basadas en la IA (véase el Recuadro 4.8).

Recuadro 4.8: Preguntas clave del Gobierno de Nueva Gales del Sur con respecto a la adopción de tecnología de IA

El gobierno de Nueva Gales del Sur utiliza tres preguntas clave para evaluar la adopción de la IA. Éstas se basan en el marco de trabajo de las “Tres V”, que fue sugerido originalmente por la empresa consultora Deloitte.

- ¿Es *viable*? Se debe entender el alcance y los límites de la tecnología y luego evaluar si la solución es viable.
- ¿Es *valioso*? El hecho de que algo pueda ser automatizado no significa que deba serlo. ¿Qué tan valiosa sería la automatización? ¿Daría algo de valor a la comunidad, y no sólo a las operaciones de su organización? ¿Cuáles serían las repercusiones? ¿Puede hacer que los resultados sean justos y éticos?
- ¿Es *vital*? ¿Su propuesta de implementación es viable sin la IA?

Fuente: www.digital.nsw.gov.au/digital-transformation/policy-lab/artificial-intelligence; www2.deloitte.com/us/en/insights/deloitte-review/issue-16/cognitive-technologies-business-applications.html.

Proporcionar perspectivas multidisciplinarias, diversas e inclusivas

Garantizar la inclusión de perspectivas multidisciplinarias (distintos antecedentes educativos, experiencias y niveles profesionales, conjuntos de aptitudes, etc.), así como diversas e inclusivas (distintos géneros, razas, edades, antecedentes socioeconómicos, etc. en un entorno en el que se valoren sus opiniones), es un factor transversal fundamental que es pertinente para las demás secciones del presente capítulo. Es quizás el principal factor que posibilita la creación de iniciativas de IA que sean a la vez eficaces y éticas, así como exitosas y justas. Es un fundamento importante para iniciativas que van desde estrategias nacionales integrales hasta pequeños proyectos individuales de IA, y todo lo demás que se pueda encontrar entre estos extremos.

La importancia de la multidisciplinariedad

El desarrollo de estrategias, proyectos y otras iniciativas de IA es un proceso intrínsecamente multidisciplinario: requiere la consideración de problemas y limitaciones tecnológicas, éticas legales y de otras restricciones. Es evidente que los esfuerzos en torno a la IA deben ser tecnológicamente viables, pero al mismo tiempo deben ser aceptables para una serie de interesados (incluido el público) y ajustarse al marco jurídico en el que se desean implementar.

Para lograrlo, las investigaciones del OPSI han demostrado que la multidisciplinariedad es uno de los factores más importantes para el éxito de los proyectos de innovación, especialmente los de carácter tecnológico. El OPSI recomienda que “al comienzo de cualquier proyecto de innovación, los gobiernos deberían convocar a un grupo formado por las personas capacitadas necesarias para que el proyecto sea un éxito. Entre ellas podrían figurar analistas y asesores de políticas, expertos de campo, diseñadores de experiencias del usuario, desarrolladores de software y abogados.”[8] Dependiendo del uso que se le dará, también podrían incluir profesiones como especialistas en sociología, psicología, medicina u otros que tengan conocimientos especializados en los ámbitos (incluidas las ciencias sociales y las humanidades) con los que una iniciativa de IA puede interactuar (Whittaker et al., 2018). Si bien su nivel de participación puede variar a lo largo del ciclo de vida de un proyecto de IA, deben tener la capacidad de mantenerse al día con respecto a los progresos y de proporcionar retroalimentación durante el proceso.

Los equipos que trabajen en la IA también tendrán que provenir de diferentes niveles de experiencia (por ejemplo, principiantes y expertos). El diseño eficaz de los servicios posibilitados por la IA a nivel operativo requerirá la competencia técnica tanto del personal de línea, como de gerentes de programas que comprendan los detalles del servicio que se está prestando y la forma en que la IA afectará al flujo de trabajo general.[9,10] Esto permitirá maximizar el impacto transformador de la IA al identificar las tareas que ya no se requieren,

[8] https://trends.oecd-opsi.org.
[9] www.pwc.com/us/en/services/consulting/library/artificial-intelligence-predictions/functional-specialists.html.
[10] Es más probable que la mayoría de los beneficios se obtengan al nivel de implementación de las políticas operativas en lugar de al nivel estratégico, aunque la IA puede tener el efecto acumulativo de facilitar nuevos enfoques estratégicos para la prestación de servicios. Para más información, véase: https://journals.sagepub.com/doi/pdf/10.1177/0952076718780537.

las nuevas tareas que son necesarias y lo que esto significa para el diseño del servicio y las aptitudes que se requieren por parte de la fuerza de trabajo (Agarwal, Gans y Goldfarb, 2018).

En el Recuadro 4.9 se presenta un ejemplo de la forma en que Data61, un componente de la agencia nacional de investigación científica de Australia (CSIRO), trata de lograr la multidisciplinariedad.

Recuadro 4.9: La multidisciplinariedad en el programa australiano Data61

Data61 se ha enfocado de forma significativa en la IA. Con la finalidad de facilitar velocidad, el grupo se formó como una “red de innovación de datos” con “límites permeables” que podrían permitir a miembros multidisciplinarios de la red con distintos antecedentes, aportar su experiencia en diversos proyectos y programas, incluidos los relativos a la IA. En una entrevista otorgada a The Mandarin, el director general de Data61, Adrian Turner, declaró que, con un modelo de esta clase, la organización “sería capaz de llevar a cabo trabajo multidisciplinario a mayor escala de una manera que no nos sería posible si se tratara únicamente de nosotros y nuestros empleados”.

Gracias a acuerdos de colaboración, Data61 ha crecido hasta convertirse en una red combinada de 1,100 personas, entre las que se encuentran expertos provenientes de 32 universidades, así como los funcionarios públicos de Data61. En la entrevista se afirmó que Turner “destaca que el talento técnico en bruto debe combinarse con la especialización en el dominio de cualquier sector en el que se apliquen, como el gobierno, la salud o la agricultura”.

Fuente: www.themandarin.com.au/112035-license-to-operate-differently-—-how-data61-achieves-large-scale-digital-innovation-with-porous-boundaries-and-multi-disciplinary-teams.

La importancia de la diversidad y la inclusión

La diversidad es un concepto global que reconoce que las personas, a pesar de ser similares en muchos aspectos, tienen experiencias de vida y características diferentes, como el género, la edad, la raza, la etnia, la capacidad física, la cultura, la religión y las creencias (Balestra y Fleischer, 2018). Estos elementos dan lugar a valores, preferencias, características y creencias únicas e importantes en cada individuo que han sido moldeados por las normas y comportamientos que han experimentado a lo largo del tiempo. La inclusión es tan importante como la diversidad en cuanto a la representación de grupos diversos, esto es, asegurar que tengan una voz y un entorno que les permita tener la oportunidad de aportar sus ideas, perspectivas y preocupaciones.

Diversas investigaciones muestran que los equipos diversos e inclusivos han demostrado ser un motor para la innovación (Forbes Insights, 2011). En el ámbito de la IA, los equipos diversos e inclusivos pueden ayudar a prevenir o eliminar posibles sesgos desde el principio (OCDE, 2019a), ya que al asegurar la representación de diferentes grupos al momento de la conceptualización y diseño del producto se ayuda a minimizar las posibilidades de sesgos en los datos y la discriminación algorítmica. También se ha demostrado en investigaciones que la diversidad en los equipos de IA genera resultados significativamente mejores para sus organizaciones (West, Kraut y Chew, 2019). Por otra parte, la falta de diversidad puede dar

lugar a algoritmos y resultados sesgados, ya que pueden reflejar cuestiones desconocidas de los grupos homogéneos que los crean (Rudgard, 2019).

Al igual que la multidisciplinariedad, la diversidad y la inclusión deben ser un aspecto importante en todo el proceso de reflexión, diseño, desarrollo y aplicación de una iniciativa de IA. Cabe señalar que, al crear aplicaciones específicas de tecnología de IA, las personas encargadas de la programación, la ingeniería y la generación de algoritmos deben provenir de orígenes diversos.

A pesar de su importancia, hay una falta de diversidad racial y de género en la investigación de la IA, la fuerza de trabajo de la IA, que refleja la falta de diversidad en el sector tecnológico y el campo de la informática en general (NSTC, 2016). Por ejemplo, apenas alrededor del 19% de las investigaciones de IA son obra de mujeres, y la proporción de mujeres que han sido coautoras de publicaciones sobre IA se ha estancado desde la década de 1990 (Owen y Stathoulopoulos, 2019). Varios gobiernos están tomando medidas para mejorar la diversidad en el campo de la IA. El Reino Unido ha sido un líder en este ámbito, como se indica en el Recuadro 4.10. En un informe reciente de la UNESCO[11] se presenta un debate interesante y exhaustivo sobre el tema de la diversidad de género en la IA en particular, con recomendaciones sobre las medidas que pueden adoptarse para mejorarla.

Recuadro 4.10: Fomentando la diversidad en la IA (Reino Unido)

Reglas para tener equipos diversos

El Foro Económico Mundial (FEM) y representantes de la Oficina de Inteligencia Artificial del Gobierno del Reino Unido y las empresas privadas Deloitte, Salesforce y Splunk diseñaron conjuntamente las directrices para la adquisición de IA. La diversidad fue un componente crucial de las directrices, y el Gobierno del Reino Unido está implementando las normas formales establecidas en las directrices que exigen la diversidad entre los equipos que trabajan en proyectos de IA para reducir el riesgo de algoritmos y sistemas de IA racistas y sexistas. Las reglas establecen que aquellos equipos que trabajan en la IA deben “incluir a personas de diferentes géneros, etnias, antecedentes socioeconómicos, discapacidades y sexualidades” y que “usted también debe asegurarse de que exista una mezcla de perspectivas y puntos de vista. De esta forma se asegura de que los problemas y las soluciones se aborden desde diferentes ángulos y ayuda a mitigar el sesgo”.

Financiamiento de los programas educativos

El Gobierno del Reino Unido ha anunciado que hará importantes inversiones financieras para promover la diversidad en el campo de la IA. Estas inversiones incluyen:

- Hasta EUR 15.7 millones (equivalente) para financiar 2,500 “títulos de conversión de especialidad universitaria” [“*conversion degrees*”] en IA y ciencia de datos, que incluyen 1,000 becas para personas provenientes de grupos subrepresentados.

11 https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1.

Recuadro 4.10: Fomentando la diversidad en la IA (Reino Unido) *(Cont.)*

- EUR 5.8 millones para impulsar la innovación en el aprendizaje en línea de los adultos a través de un Fondo de Innovación Tecnológica para el Aprendizaje de los Adultos [*Adult Learning Technology Innovation Fund*].

Los esfuerzos están destinados a impulsar la diversidad de género y a mejorar la capacitación y el readiestramiento de los adultos para que adquieran aptitudes en torno a las tecnologías modernas, a fin de ayudarles a progresar en nuevos aspectos de sus carreras actuales o a encontrar nuevas oportunidades en ámbitos relacionados con la tecnología. El financiamiento de la educación en línea "proporcionará fondos y conocimientos especializados destinados a estimular a las empresas de tecnología para que hagan uso de las nuevas tecnologías, y así creen oportunidades de capacitación en línea a la medida, flexibles, inclusivas y atractivas que permitan a más personas participar en empleos especializados".

Fuente: www.telegraph.co.uk/technology/2019/09/20/government-ai-rules-require-diverse-teams-prevent-racist-sexist; www3.weforum.org/docs/WEF_Guidelines_for_AI_Procurement.pdf; www.digital.nsw.gov.au/digital-transformation/policy-lab/artificial-intelligence; www.gov.uk/government/news/185-million-to-boost-diversity-in-ai-tech-roles-and-innovation-in-online-training-for-adults.

Desarrollar una estrategia fiable, justa y responsable

Hay una variedad de factores que intervienen en el desarrollo de una estrategia fiable, justa y responsable. Estos se analizan en las subsecciones siguientes.

Establecer marcos jurídicos, éticos y técnicos en la etapa de diseño y vigilar su cumplimiento durante la fase de ejecución

La Inteligencia Artificial puede tener un efecto transformador y disruptivo en la forma en que se prestan los servicios públicos y en el desempeño de las administraciones. Esta disrupción significa que las trayectorias de la IA están definidas por la complejidad, la incertidumbre y el riesgo (Mateos-García, 2018). De tal forma, la elaboración de marcos rigurosos para configurar la toma de decisiones en los organismos del sector público será crucial para hacer realidad el potencial de la IA de transformar los servicios públicos y la administración. Aunque es importante evaluar e iterar los proyectos e iniciativas de IA a lo largo de su ciclo de vida, resulta vital garantizar que el diseño sea ético e imparcial desde las primeras etapas, ya que será más costoso atender los problemas más adelante durante la implementación.

Como se señaló anteriormente, la articulación de principios claros para la IA ayuda a crear un entorno propicio que se ajusta en general a los objetivos y valores sociales de dichos principios. Es probable que el compromiso abierto con los principios éticos sea una condición necesaria pero no suficiente para el despliegue efectivo de la IA. Para que los principios tengan el máximo impacto en las conductas, tendrán que ser viables e incorporarse a los procesos e instituciones que dan forma a la toma de decisiones dentro del gobierno. Los investigadores del Instituto del Internet de Oxford [*Oxford Internet Institute*][12]

[12] www.oii.ox.ac.uk.

y del Instituto Alan Turing [*Alan Turing Institute*] han colaborado con otras instituciones para sintetizar los principios éticos, así como los factores subyacentes y las mejor prácticas correspondientes para la IA, a fin de ayudar a actualizarlos (véase el Recuadro 4.11).

Recuadro 4.11: Un marco ético para una buena sociedad de la IA (AI4People) y la IA para el Bien Social (AI4SG)

Un artículo reciente publicado por AI4People ofrece una síntesis de las expresiones existentes de los principios éticos para la IA generadas por organizaciones de renombre. El artículo resume varios temas clave de estos principios:

- **Beneficencia** – promover el bienestar, preservar la dignidad y sostener el planeta
- **No maleficencia**- privacidad, seguridad y "precaución respecto de las capacidades" (no hacer daño y evitar el mal uso/sobreutilización de la tecnología)
- **Autonomía** - "es la idea de que las personas tienen derecho a tomar decisiones por sí mismas con respecto al tratamiento que reciben o dejan de recibir"
- **Justicia** - promover la prosperidad y preservar la solidaridad (equidad, no discriminación y asegurar que los beneficios se compartan ampliamente).
- **Explicabilidad** - posibilitar los demás principios a través de la inteligibilidad y la responsabilidad.

En un documento aparte, analizan una serie de casos de estudio para exponer los factores esenciales que sirven de base para el diseño de sistemas de IA para el Bienestar Social (AI4SG) exitosos.

Factores	Las mejores prácticas correspondientes
Refutabilidad y despliegue incremental	Identificar los requisitos refutables y probarlos en pasos incrementales desde el laboratorio hasta el "mundo exterior".
Salvaguardas contra la manipulación de las variables predictivas	Adoptar salvaguardas que: (i) garanticen que los indicadores no causales no desvíen indebidamente las intervenciones; y (ii) limiten, según corresponda, el conocimiento de la forma en que los datos afectan los resultados de los sistemas de AI4SG, con la finalidad de evitar la manipulación.
Intervención contextualizada en torno al receptor	Crear sistemas de toma de decisiones con aportes de los usuarios que interactúan con esos sistemas y que se ven afectados por ellos, incluyendo el entendimiento de las características de los usuarios, los métodos de coordinación, los fines y los efectos de una intervención, y respetando el derecho de los usuarios a ignorar o modificar las intervenciones.
Explicación contextualizada en torno al receptor y propósitos transparentes	Elegir un Nivel de Abstracción para la explicación de la IA que cumpla con el propósito explicativo deseado y que sea adecuado para el sistema y los receptores, luego exponer aquellos argumentos que sean fundamentalmente persuasivos y suficientes para que el receptor dé la explicación, y garantizar que el objetivo (el fin del sistema) para el cual se desarrolla y despliega un sistema AI4SG sea conocido por los receptores de sus resultados de forma predeterminada.
Protección de la privacidad y consentimiento del propietario de los datos	Respetar el umbral de consentimiento establecido para el procesamiento de los conjuntos de datos personales.

Recuadro 4.11: Un marco ético para una buena sociedad de la IA (AI4People) y la IA para el Bien Social (AI4SG) *(Cont.)*

Factores	Las mejores prácticas correspondientes
Igualdad de circunstancias	Eliminar de los conjuntos de datos pertinentes aquellas variables y valores representativos que sean irrelevantes para un resultado, excepto cuando su inclusión respalde la inclusividad, la seguridad u otros imperativos éticos.
Semantización amigable	No obstaculizar la capacidad de las personas para semantizar (es decir, dar sentido y significado a) algo.

Fuente: https://link.springer.com/article/10.1007/s11023-018-9482-5; www.research.ed.ac.uk/portal/files/77587861/FloridiEtalMM2018AI4PeopleAnEthicalFrameworkFor.pdf.

La Directiva sobre la Toma de Decisiones Automatizada [*Directive on Automated Decision-Making*] del Gobierno de Canadá tiene por objeto poner en funcionamiento un conjunto de principios jurídicos, éticos y técnicos diseñados para garantizar la existencia de normas y un enfoque uniforme sobre la gestión de riesgos en materia de IA en todo el sector público, tanto en la fase de diseño como en la de implementación. Como acompañamiento a la Directiva, el Gobierno de Canadá elaboró una Evaluación de Impacto Algorítmico [*Algorithmic Impact Assessment*] que valora el posible impacto de un algoritmo sobre los ciudadanos. De esta forma se obtienen medidas altamente detalladas, basadas en riesgos, que les permiten a los funcionarios centrarse en implementar una mitigación eficaz cuando los riesgos son mayores. La Directiva y la Evaluación se examinan a fondo en un caso de estudio que figura en el Anexo A. El caso de estudio sobre las Directrices Éticas para una IA fiable de la Comisión Europea también ofrece información a considerar para evaluar las cuestiones éticas.

Se necesitará vigilancia durante la etapa de implementación para garantizar que el sistema funcione según lo previsto, que se mitiguen los riesgos y que se identifiquen las consecuencias indeseadas. Se requerirá un enfoque diferenciado para centrar la atención en los sistemas de IA en los lugares donde los riesgos son mayores, por ejemplo, donde influyen en la distribución de los recursos o tienen otras consecuencias importantes para los ciudadanos (Mateos-Garcia, 2017). En el Recuadro 4.12 se expone la estrategia de la ciudad de Nueva York para supervisar su uso de la IA.

Recuadro 4.12: El Grupo de trabajo sobre sistemas de toma de decisiones automatizadas de la ciudad de Nueva York

El alcalde de Nueva York anunció la creación de un grupo de trabajo para supervisar el uso de la IA en la ciudad con el fin de garantizar la responsabilidad, la equidad y la justicia en todas sus áreas de responsabilidad. A más tardar en diciembre de 2019, el grupo recomendará procedimientos destinados a examinar y evaluar los instrumentos de IA a fin de garantizar la equidad y la existencia de oportunidades. El objetivo es promover la transparencia y el apego congruente a normas y valores comunes. El grupo de trabajo estará integrado por funcionarios responsables de los servicios, académicos, expertos en derecho y tecnología, grupos de la sociedad civil y grupos de reflexión.

Fuente: www1.nyc.gov/office-of-the-mayor/news/251-18/mayor-de-blasio-first-in-nation-task-force-examine-automated-decision-systems-used-by.

Mantener una atención constante en los usuarios y en quienes pueden verse afectados

En la sección anterior, titulada "¿Es la IA la mejor solución al problema?", se examinó la importancia de comprender las necesidades de los usuarios al momento de considerar las posibles iniciativas y alternativas de la IA. Sin embargo, este enfoque en los usuarios no termina cuando se descubren cuáles son las necesidades inmediatas, ni sus objetivos se limitan únicamente a comprender cuál solución tecnológica es la adecuada. Los gobiernos deben seguir colaborando con los usuarios y otras personas que puedan verse afectadas directa o indirectamente durante todo el ciclo de vida de la solución de IA, a medida que se implementa, evalúa e itera. Como se muestra en el Recuadro 4.11, este proceso a menudo se expresa mediante el desarrollo de principios éticos.

En 2018, el Instituto Alan Turing publicó una guía para el diseño y la implementación responsables de la IA en el sector público (Leslie, 2019). En esta guía se afirma que los gobiernos deben empezar por "construir en forma ascendente desde la base cultural" a fin de establecer una plataforma ética para la IA. En su nivel más básico, esto requiere comprender los valores que "*Respalden, Protejan y Motiven*" ["*Support, Underwrite and Motivate*"] (Valores SUM, por sus siglas en inglés) un ecosistema de IA responsable. Estos valores se apegan al principio de que se considere y consulte durante el ciclo de vida de la IA a todos los usuarios y otras personas que puedan verse afectadas. Los valores SUM consisten en *respetar, conectar, cuidar y proteger*, como se detalla en el Recuadro 4.13.

Recuadro 4.13: Los valores SUM: Respetar, conectar, cuidar y proteger

Los valores SUM están "pensados para utilizarse como valores rectores a lo largo del ciclo de vida de la innovación: desde las etapas preliminares de evaluación, planificación y formulación de problemas de los proyectos, pasando por los procesos de diseño, desarrollo y pruebas, hasta las etapas de implementación y reevaluación".



***Respetar* la dignidad de las personas:**

- Salvaguardar su autonomía, su poder de expresión y su derecho a ser escuchados.
- Apoyar su posibilidad de prosperar, desarrollarse plenamente, dedicarse a lo que les apasiona y desarrollar sus talentos de acuerdo con sus propios planes de vida que ellos mismos han determinado.

Recuadro 4.13: Los valores SUM: Respetar, conectar, cuidar y proteger *(Cont.)*

***Conectar* con los demás de forma sincera, abierta e inclusiva:**

- Dar prioridad a la diversidad, la participación y la inclusión en todos los puntos de los procesos de diseño, desarrollo y despliegue de innovaciones de IA.
- Exhortar a que todas las voces sean escuchadas y que todas las opiniones sean tomadas en cuenta seria y honestamente durante todo el ciclo de vida de producción y uso.

***Cuidar* del bienestar de todos:**

- Diseñar y desplegar sistemas de IA para fomentar y cultivar el bienestar de todas las partes interesadas cuyos intereses se vean afectados por su uso.
- Dar prioridad a la seguridad e integridad mental y física de las personas cuando se exploran los horizontes de posibilidades tecnológicas y al concebir y desplegar aplicaciones de IA.

***Proteger* las prioridades de los valores sociales, la justicia y el interés público:**

- Dar prioridad al bienestar social, al interés público y a la consideración de las repercusiones sociales y éticas de las innovaciones al momento de determinar la legitimidad y la conveniencia de las tecnologías de IA.
- Reflexionar detenidamente sobre los impactos más amplios de las tecnologías de IA que se están concibiendo y desarrollando. Considerar las ramificaciones que tendrán sus efectos y otros factores externos para otros alrededor del mundo, para las generaciones futuras y para la biosfera como conjunto.

Fuente: www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf.

Actualmente existe una serie de herramientas y recursos potenciales que los funcionarios gubernamentales pueden utilizar para ayudar a identificar y colaborar con los usuarios y las personas que pueden verse afectadas por un sistema de IA, a fin de comprender mejor su perspectiva. El Navegador de Herramientas [*Toolkit Navigator*] del OPSI[13] puede ser de ayuda al momento de navegar a través de una vasta colección de herramientas de innovación y poder encontrar las que mejor se adapten a sus necesidades en función de las circunstancias. Entre los ejemplos de estas herramientas se incluyen:

- **Tarjetas metodológicas de 18F [18F *Method Cards*]:** Un juego de tarjetas imprimibles que incluye información simplificada sobre diversos métodos de diseño centrados en el ser humano según una metodología general (Descubrir, Decidir, Hacer y Validar)[14]
- **Menú de métodos de diseño social [*Social Design Methods Menu*]:**[15] Proporciona un enfoque y métodos para abordar problemas de índole social y política al hacer que los servicios sean más valiosos y fáciles de usar para los clientes y usuarios, ya que se

13 https://oecd-opsi.org/toolkit-navigator.
14 https://methods.18f.gov.
15 www.lucykimbell.com/stuff/Fieldstudio_SocialDesignMethodsMenu.pdf.

desperdician menos recursos en la implementación de ideas equivocadas o de ideas correctas de manera equivocada. Este enfoque implica que se dedique tiempo a comprender las experiencias y los recursos de las personas en sus propios términos, a adoptar pasos metódicos para analizar y abordar los problemas con su participación activa, e impulsar un trabajo más eficaz entre los equipos y las organizaciones.

- **Guía de campo del kit de diseño de IDEO para el diseño centrado en el ser humano [*IDEO Design Kit Field Guide to Huma-Centered Design*]:**[16] Ofrece orientación sobre las distintas fases de los procesos de diseño centrados en el ser humano, con una clasificación por mentalidades, métodos (inspiración, conceptualización e implementación) y herramientas.

Aclarar el papel adecuado de los humanos en el proceso de toma de decisiones

En muchos casos, si no es que, en todos, se recomienda a los gobiernos permitir que los humanos participen de forma activa en este proceso, particularmente cuando se está desplegando un nuevo sistema. En esos casos, será fundamental que los funcionarios que trabajan junto con el sistema de IA tengan clara su función precisa en el proceso de toma de decisiones. En el Recuadro 4.14 se presenta un marco para reflexionar sobre los distintos niveles de interacción entre hombre y máquina. Los funcionarios deberán poseer los conocimientos y aptitudes necesarios para comprender el funcionamiento del sistema de IA, así como sus fortalezas y debilidades, de tal manera que puedan monitorearlo de forma eficaz y detectar anomalías. También deben tener la certeza de su propio nivel de autoridad para tomar decisiones. Para que los resultados sean eficaces, también será necesario que los sistemas sean técnicamente sólidos y seguros (Comisión Europea, 2019a).

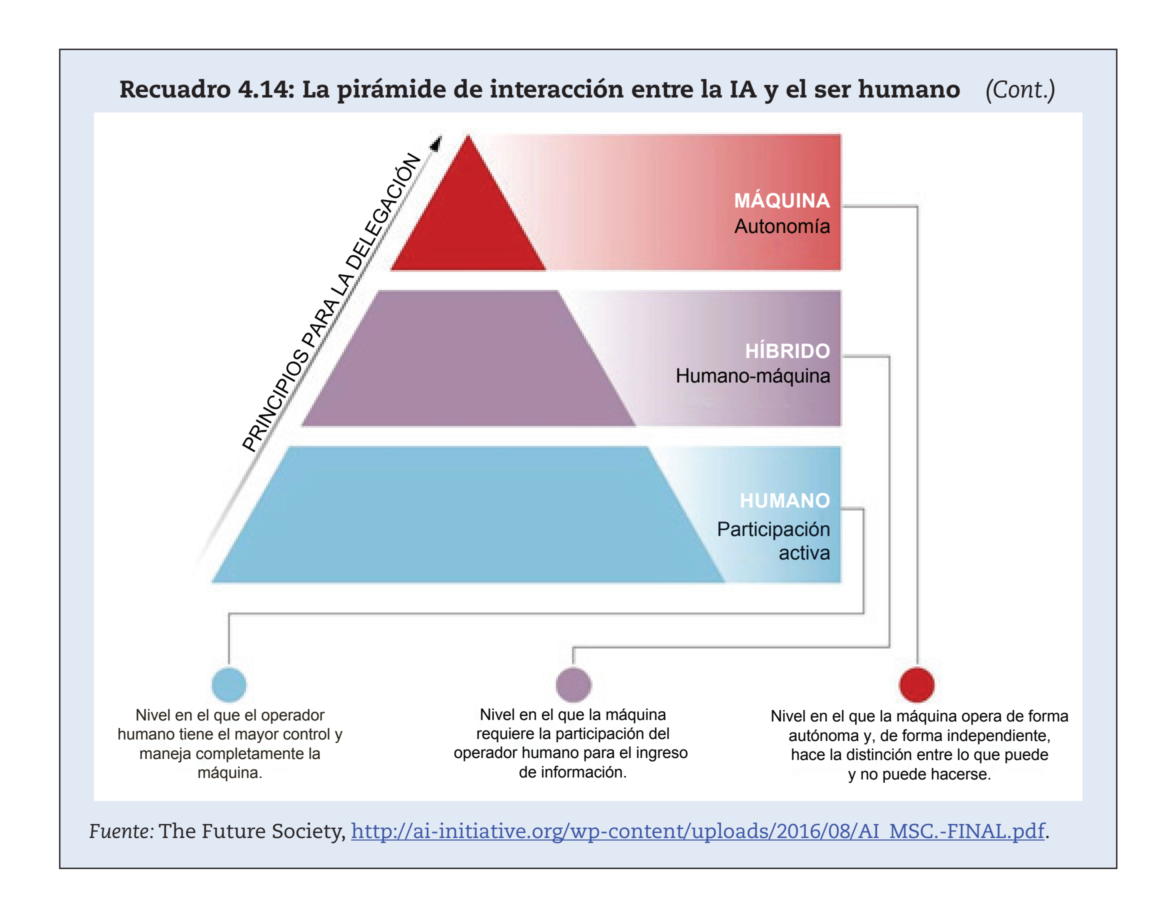
Recuadro 4.14: La pirámide de interacción entre la IA y el ser humano

La Iniciativa para la IA en la Sociedad Futura [*The Future Society*] (TFS, por sus siglas en inglés), incubada en la Escuela de Gobierno Kennedy de Harvard [*Harvard Kennedy School of Government*] (HKS, por sus siglas en inglés), elaboró un marco sencillo para comprender la esencia de la interacción entre la IA y los seres humanos en el contexto de los conflictos armados.

La Pirámide de Interacción IA-ser humano les permite a los dirigentes del sector público evaluar la esencia de dichas interacciones en los sistemas de IA de los que son responsables. Esto constituye un primer paso para determinar si son adecuadas dados los costos, beneficios y riesgos asociados.

[16] www.designkit.org.

Recuadro 4.14: La pirámide de interacción entre la IA y el ser humano *(Cont.)*



Fuente: The Future Society, http://ai-initiative.org/wp-content/uploads/2016/08/AI_MSC.-FINAL.pdf.

Es importante que los funcionarios públicos comprendan que el hecho de contar con controles eficaces no reducirá el riesgo a cero. Es necesario entrenar a los algoritmos para que puedan prestar un servicio viable, y la obligación de experimentar sugiere que siempre existe la posibilidad de que una IA no funcione de la forma esperada. Incluso, aunque se trate de un algoritmo imparcial, es poco probable que éste sea preciso en un 100%. Sin embargo, también es importante considerar lo contrario: posponer el despliegue de la IA significará un retraso en la obtención de los beneficios que ésta puede aportar, y es poco probable que los procesos de toma de decisiones existentes sean completamente exactos e imparciales. Por consiguiente, los gobiernos tendrán que determinar cuál es el equilibrio adecuado entre los controles estrictos y la experimentación y el riesgo, con base en los costos y beneficios relativos.

Desarrollar estructuras de responsabilidad abiertas y transparentes

Es probable que establecer procesos y estructuras fiables, justos y responsables ayude a los gobiernos a alcanzar el potencial de la IA para transformar los servicios públicos y la administración, así como a fomentar la confianza del público en su capacidad para hacerlo. Si el público no confía en que el gobierno utilice la IA de forma ética, puede evitar los servicios que utilizan la IA u oponerse a su introducción. Por lo tanto, será importante abordar las inquietudes del público, y esto puede respaldarse mediante una supervisión rigurosa, la rendición de cuentas y los procesos justos.

La transparencia y la rendición de cuentas dependen de la adopción de marcos jurídicos, éticos y técnicos, así como de sistemas de vigilancia de la implementación y para la

gestión de los riesgos, como se ha señalado anteriormente. Los códigos para la práctica y las decisiones/reglas ayudan a determinar cuáles son las circunstancias en las se permite el uso de la IA en el sector público y qué controles y salvaguardas será necesario establecer. La adopción uniforme de esos marcos puede contribuir a promover la equidad procesal, el cumplimiento de la ley y el debido proceso. Sin embargo, sólo contribuirán a la rendición de cuentas si se comunican al público de forma clara y sencilla. Por ejemplo, el Gobierno de Canadá exige a los organismos del sector público que publiquen los resultados de su Evaluación de Impacto Algorítmico en forma de datos abiertos al público, para contribuir a la concientización de la población respecto a las decisiones que pudieran afectarle (véase el caso de estudio del Anexo A).

Es más probable que los marcos de responsabilidad sean eficaces si los gobiernos brindan suficiente información sobre sus actividades de IA para facilitar el escrutinio por parte de los interesados externos, incluidos los expertos. Por ejemplo, el Centro de Ética e Innovación de Datos [*Centre for Data Ethics and Innovation*] del Gobierno del Reino Unido (véase el Recuadro 4.26 más adelante en este capítulo) vigila el uso de la IA en el sector público a través de evaluaciones de administración que permitan identificar las brechas, los riesgos y las oportunidades, así como a recomendar mejoras.[17] En otro ejemplo, Etalab, el grupo de trabajo del Primer Ministro de Francia para datos abiertos y gobierno abierto, publicó una guía para las administraciones públicas sobre la forma en que deben utilizarse los algoritmos, y hace hincapié en la transparencia y la rendición de cuentas (véase el Recuadro 4.15). A nivel mundial, la Asociación de Maquinaria de Computación [*Association for Computing Machinery*] (ACM, por sus siglas en inglés) ha celebrado desde 2018 una conferencia anual sobre Equidad, Responsabilidad y Transparencia (FAT, por sus siglas en inglés), que reúne a académicos de diversos ámbitos, como la informática, el derecho, las ciencias sociales y las humanidades, para examinar cuestiones conexas que a menudo implican la IA y el Aprendizaje Automático. La Red ACM FAT* es una serie de talleres sobre los mismos temas.[18]

Recuadro 4.15: Guía de Etalab sobre la responsabilidad de los algoritmos públicos (Francia)

Etalab, el Grupo de Trabajo dependiente de la Oficina del Primer Ministro francés y el encargado de los datos abiertos y el gobierno abierto, ha elaborado una guía para las administraciones públicas sobre el uso responsable de los algoritmos en el sector público. La guía establece la forma en que los organismos deben informar sobre su uso para promover la transparencia y la rendición de cuentas.

Esta guía forma parte de un programa de trabajo sobre algoritmos públicos que también incluye la elaboración de casos de estudio, la identificación de proyectos de IA en el sector público y el apoyo técnico a dichos proyectos, la anticipación del impacto de la IA en las partes interesadas y la reflexión sobre cuestiones éticas relacionadas con el uso de la IA en la esfera pública.

17 www.gov.uk/government/groups/centre-for-data-ethics-and-innovation-cdei.
18 https://fatconference.org.

Recuadro 4.15: Guía de Etalab sobre la responsabilidad de los algoritmos públicos (Francia) *(Cont.)*

Ésta abarca tres elementos:

- **Elementos contextuales.** Se centran en la naturaleza de los algoritmos, la forma en que pueden utilizarse en el sector público y la distinción entre las decisiones automatizadas y los casos en que los algoritmos funcionan como instrumentos de apoyo a la toma de decisiones.
- **La ética y la responsabilidad de usar algoritmos para mejorar la transparencia**. Incluye la presentación de informes públicos sobre el uso de los algoritmos, la forma de garantizar una toma de decisiones justa e imparcial, y la importancia de la transparencia, la explicabilidad y la fiabilidad.
- **El marco jurídico para la transparencia en los algoritmos**, incluido el Reglamento General de Protección de Datos de la Unión Europea (RGPD) y la legislación nacional. Esto incluye un conjunto de normas que se aplicarán a los procesos de toma de decisiones administrativas sobre qué información específica debe publicarse respecto a los algoritmos públicos.

Etalab también propone seis principios rectores para la responsabilidad de la IA en el sector público:

- **Reconocimiento**: las dependencias están obligadas a informar a las partes interesadas cuando se utiliza un algoritmo.
- **Explicación general**: las dependencias deben dar una explicación clara y comprensible de cómo funciona un algoritmo.
- **Explicación individual**: las dependencias deben dar una explicación personalizada de un resultado o decisión específicos.
- **Justificación**: las dependencias deben justificar los motivos por los que se utiliza un algoritmo y las razones para elegir un algoritmo determinado.
- **Publicación**: las dependencias deben publicar el código fuente y la documentación, además de informar a las partes interesadas si el algoritmo fue creado o no por un tercero.
- **Permitir la impugnación**: las dependencias deben ofrecer formas de debatir sobre los procesos algorítmicos e impugnarlos.

Fuente: www.etalab.gouv.fr/datasciences-et-intelligence-artificielle; www.etalab.gouv.fr/how-etalab-is-working-towards-public-sector-algorithms-accountability-a-working-paper-for-rightscon-2019; https://github.com/etalab/algorithmes-publics/blob/master/20190611_WorkingPaper_PSAAccountability_Etalab.pdf; www.europeandataportal.eu/fr/news/enhancing-transparency-through-open-data; www.etalab.gouv.fr/algorithmes-publics-etalab-publie-un-guide-a-lusage-des-administrations.

Considerar la explicabilidad de los sistemas de IA y la toma de decisiones automatizada

Para que la responsabilidad funcione eficazmente, los gobiernos deben ser capaces de explicar por qué un sistema de IA tomó determinadas decisiones, en particular si la decisión tiene el potencial de influir en la vida de las personas. No obstante, la complejidad de los algoritmos de IA podría dificultar que se pueda dar una explicación clara que justifique una decisión. La IA busca realizar predicciones o inferencias óptimas basadas en correlaciones; no depende de una teoría o historia global que explique por qué esas correlaciones son relaciones causales importantes (Anastasopoulos y Whitford, 2019). Además, es probable que las leyes de protección de datos incluyan una exigencia en torno a la explicabilidad. En general se ha informado que, conforme a las disposiciones del RGPD, los organismos están obligados a explicar a los ciudadanos cómo la IA utiliza sus datos para tomar decisiones automatizadas (Raja, 2018). La cuestión de si eso sucede en realidad, y si es incluso factible, sigue en debate (Wachter, Mittelstadt y Floridi, 2017). Dado que este debate podría llevar algo de tiempo para resolverse, los gobiernos deben tratar de explicar tales decisiones.

Independientemente de la forma en que se tomen las decisiones, es importante saber con precisión quién está facultado para tomar decisiones sobre la forma en que se despliega la IA, quién es responsable de cada decisión y a quién debe rendir cuentas. Evidentemente, estos procesos de rendición de cuentas son cruciales cuando las decisiones podrían tener un impacto significativo en la vida de las personas. Los marcos de gobernanza que otorguen voz y facultades de vigilancia a los usuarios de los servicios serán especialmente importantes (Whittaker, 2018). A este respecto, existen diversas aproximaciones que los gobiernos podrían adoptar para mitigar los problemas de explicabilidad y posibilitar la rendición de cuentas:

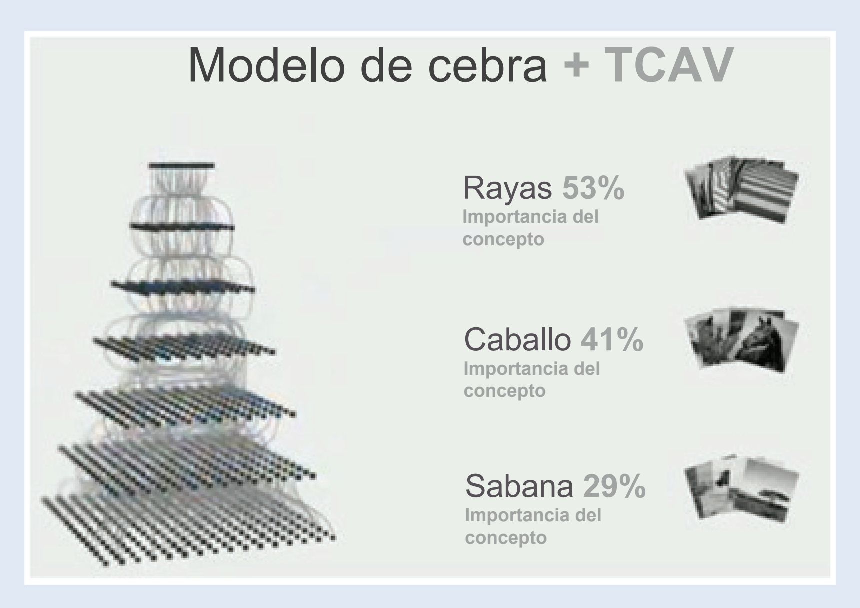
- **Crear una IA explicable**. Los gobiernos podrían emprender esfuerzos para crear una IA que sea explicable por naturaleza. Sin embargo, esto puede dar lugar a un aumento del costo para favorecer la interpretabilidad.[19] Ofrecer explicaciones como las que esperan las personas en virtud de la legislación vigente "con frecuencia debería ser técnicamente factible, pero en ocasiones esto puede ser prácticamente oneroso" (Kortz y Doshi-Velez, 2017; Recuadro 4.16).

- **Enfoques con participación humana activa** Los casos en los que la IA sirve como apoyo a la toma de decisiones por parte de los funcionarios públicos, podrían plantear menos problemas de explicabilidad que la toma de decisiones totalmente automatizada, pero no mitigarán totalmente el riesgo y aumentarán los costos de los sistemas de IA, especialmente cuando se despliegan a una mayor escala (Mateos-Garcia, 2017). Si los resultados del algoritmo afectan las decisiones de los funcionarios, entonces es necesario que tanto ellos como el público tengan la capacidad de entender los motivos por los que el algoritmo recomendó esa decisión.

[19] www.pwc.com/us/en/services/consulting/library/artificial-intelligence-predictions/explainable-ai.html.

Recuadro 4.16: Creación de redes neuronales explicables con Pruebas con Vectores de Activación de Concepto (TCAV, por sus siglas en inglés)

Las redes neuronales tienen el potencial de realizar predicciones muy precisas; sin embargo, su complejidad hace que sean difíciles de explicar. No obstante, se están realizando esfuerzos para explicar incluso la IA más compleja. Por ejemplo, Google está explorando las Pruebas con Vectores de Activación de Concepto (TCAV) para determinar cuáles son las señales que utilizan las redes neuronales para realizar sus predicciones. Esto permitirá identificar cuáles son los factores más destacados para determinar una decisión, incluidas las fuentes de sesgo. El siguiente ejemplo ilustra los conceptos que podrían ser importantes para que un algoritmo identifique la imagen de una cebra:



Fuente: www.zdnet.com/article/google-says-it-will-address-ai-machine-learning-model-bias-with-technology-called-tcav.

Establecer salvaguardas contra los sesgos y la injusticia

Un objetivo clave de la transparencia de la IA consiste en mitigar y controlar el sesgo y la justicia distributiva en la toma de decisiones algorítmica. Si las decisiones son tomadas por un sistema de "caja negra" de aprendizaje profundo, será más difícil vigilar si los resultados contienen sesgos y saber si podrían tener un efecto injusto en la vida de las personas. Por consiguiente, es necesario crear marcos de gobernanza en la etapa de diseño que incluyan un medio de supervisar los resultados para identificar y reducir la discriminación por características como el origen étnico, el género, los ingresos, la situación de discapacidad y la edad. Si no se cuenta con los medios para limitar los sesgos de una IA, será complicado justificar su uso en el sector público. En el Recuadro 4.17, se exponen las cuestiones relativas a los sesgos en las evaluaciones de los riesgos de justicia penal en los Estados Unidos.

Recuadro 4.17: Inquietudes sobre el sesgo algorítmico en el sistema de justicia penal de EE. UU.

En algunas partes del sistema de justicia penal de los Estados Unidos, los jueces utilizan evaluaciones de riesgos que valoran la probabilidad de que un delincuente vuelva a delinquir para fundamentar las decisiones relativas a las sentencias, el acceso a los servicios de rehabilitación y para decidir si un acusado permanecerá detenido mientras espera su juicio.

En teoría, las decisiones basadas en datos deberían reducir el sesgo en las decisiones de los jueces. Sin embargo, los algoritmos estiman las tasas de reincidencia basándose en las correlaciones históricas entre las variables, las cuales no representan necesariamente relaciones causales. Por lo tanto, si estas correlaciones se ven afectadas por el sesgo, entonces la discriminación se arraigará en el sistema. Por ejemplo, si las decisiones de los jueces anteriores se han visto afectadas por un sesgo, o si existe una correlación entre, por ejemplo, el origen étnico o los ingresos y la reincidencia, entonces las personas podrían recibir sentencias más severas debido a estas características. Por lo tanto, “el algoritmo podría amplificar y perpetuar los sesgos que ya se han arraigado y generar aún más datos sesgados para alimentar un círculo vicioso”.

El análisis de los resultados de un modelo de evaluación de riesgos utilizado en el condado de Broward (Florida) reveló que, al comparar la reincidencia prevista con las tasas de reincidencia reales, a menudo se anticipaba que los acusados de raza negra presentaban un riesgo de reincidencia mayor que el real, mientras que a los acusados blancos se les consideraba con frecuencia con un riesgo menor. Además, dado que los algoritmos son software de propiedad exclusiva, no siempre es posible acceder al código fuente para comprender la forma en que se toman las decisiones.

Un informe de la coalición Partnership for AI identificó tres conjuntos de problemas derivados del uso de estas evaluaciones de riesgos:

- **Inquietudes sobre la precisión, el sesgo y la validez de las herramientas mismas**: no debe presumirse que las herramientas son objetivas e imparciales simplemente porque se basan en datos.
- **Problemas con la interfaz que facilita la interacción entre las herramientas y los humanos que las utilizan**: debe ser posible interpretar y explicar las herramientas de tal forma que los usuarios puedan entender la forma en que éstas realizan predicciones.
- **Cuestiones de administración, transparencia y responsabilidad**: estas predicciones tienen un impacto significativo en la vida de los ciudadanos, por lo que las personas que “especifican, ordenan y despliegan” estas herramientas deben responder por lo anterior.

En consecuencia, la coalición recomienda que se evite el uso de herramientas de evaluación de riesgos o que se establezcan normas para mitigar cada una de esas cuestiones.

Fuente: www.technologyreview.com/s/612775/algorithms-criminal-justice-ai; www.partnershiponai.org/artificial-intelligence-research-and-ethics-community-calls-for-standards-in-criminal-justice-risk-assessment-tools; www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm, https://ainowinstitute.org/AI_Now_2018_Report.pdf.

La existencia de sesgos en los datos de contratación de personal es otro ejemplo, ya que los algoritmos creados con datos sesgados pueden perpetuar estos sesgos. El ejemplo más conocido de esto fue la herramienta experimental de aprendizaje automático de Amazon utilizada para calificar a los candidatos a un puesto de trabajo. Se demostró que esta herramienta tenía un sesgo contra las mujeres porque se entrenó con datos con sesgos incorporados que reflejaban un campo dominado por los hombres. "En efecto, el sistema de Amazon se enseñó a sí mismo que los candidatos masculinos eran preferibles. Penalizó los currículums que incluían la palabra "femenino", como en "capitana del club de ajedrez femenino". Y bajó la calificación de las graduadas de dos universidades exclusivamente para mujeres" (Dastin, 2018).

Sin embargo, no debe asumirse que el sesgo de la IA es una barrera inevitable. Mejorar la alimentación de los datos, realizar ajustes para atender el sesgo y eliminar las variables que causan el sesgo podría hacer que las aplicaciones de IA sean más justas y precisas. Por ejemplo, para contrarrestar el sesgo en las contrataciones, el Gobierno del Reino Unido ayudó a proporcionar el financiamiento inicial y posteriormente se asoció con Be Applied, una plataforma de contratación que utiliza la ciencia del comportamiento para eliminar el sesgo inconsciente del proceso de contratación mediante la anonimización y la aleatorización de los datos.[20] Como vimos anteriormente en este capítulo, la formación de equipos diversos y la incorporación de la revisión por pares también mitigarán el sesgo (Moneycontrol News, 2019). En muchos casos, las decisiones automatizadas pueden tener el potencial de ser más justas que las decisiones humanas, si únicamente toman en cuenta la información pertinente y lo hacen de manera transparente y explicable.[21]

Además del sesgo, también existen problemas de equidad en la distribución de los servicios y el estigma social relacionado con el uso de la IA. Las "puntuaciones de datos" que combinan datos de diversas fuentes como forma de clasificar a los ciudadanos, asignar servicios y predecir el comportamiento se han vuelto cada vez más comunes en los servicios públicos. Ese tipo de puntuación puede utilizarse con fines cuestionables que pueden dar lugar a un mayor arraigo de las desigualdades sociales. El ejemplo de las puntuaciones del Crédito Social de China (Recuadro 4.18) ilustra algunos de los problemas relacionados con las puntuaciones de datos.

Recuadro 4.18: Las puntuaciones del Crédito Social de China

En varias ciudades chinas se están llevando a cabo ensayos de un sistema de crédito social que puede influir en el acceso a servicios, créditos, empleos y viajes en función de si se considera que el ciudadano es digno de confianza. El sistema que determina la puntuación de crédito social funciona con sistemas de IA, que incluyen la tecnología de reconocimiento facial ligada a la vigilancia mediante sistemas de CCTV, la recopilación de datos a través de aplicaciones de teléfonos inteligentes para medir el comportamiento en línea, los activos financieros y registros gubernamentales, como las evaluaciones educativas, médicas y de seguridad del Estado.

20 https://oecd-opsi.org/innovations/applied-using-behavioural-science-to-remove-unconscious-bias-from-recruitment.

21 www.digital.nsw.gov.au/digital-transformation/policy-lab/artificial-intelligence.

Recuadro 4.18: Las puntuaciones del Crédito Social de China *(Cont.)*

Esto les brinda a las autoridades la capacidad de controlar y moldear el comportamiento de los ciudadanos. Todo lo que alguien diga, compre y con quién se asocia puede influir en su capacidad de participar en la vida pública. Esto puede tener un efecto intimidatorio con respecto al desacuerdo y la supervisión rigurosa del Estado por parte de la población.

Parece probable que este tipo de sistema de crédito social sea tecnológicamente factible en muchos países mediante la agregación de los datos de las personas procedentes de diversas fuentes, pero eso no quiere decir que sea deseable o inevitable. El surgimiento de estos sistemas, así como los controles a los que están sujetos, son cuestiones políticas importantes. Las respuestas podrían depender en parte del equilibrio que se dé a la importancia de una sociedad estable y segura o a la privacidad y la libertad individual.

Fuente: www.abc.net.au/news/2018-09-18/china-social-credit-a-model-citizen-in-a-digital-dictatorship/10200278; https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf; https://time.com/collection/davos-2019/5502592/china-social-credit-score.

Sin embargo, los gobiernos a menudo utilizan prácticas similares para atender problemas sociales apremiantes. Por ejemplo, ante las altas tasas de refugiados en busca de mejores condiciones de vida, Suiza utiliza un programa piloto de utilización de perfiles de refugiados basados en datos, los cuales se analizan mediante algoritmos, para situar a los refugiados en las zonas en las que tendrán más posibilidades de integrarse con éxito, incluidas las posibilidades de empleo. Se cree que el algoritmo aumentará los resultados de empleo en un 40-70% en promedio en comparación con las cifras actuales (Bansak et al., 2018).

En el Reino Unido, los gobiernos locales y las fuerzas policiales han tratado en algunos casos de combinar una variedad de conjuntos de datos, por ejemplo, para utilizar la IA con el fin de predecir cuáles son los niños que corren el riesgo de ser víctimas de abuso o abandono, a fin de orientar mejor los servicios (Dencik et al., 2018) o para identificar patrones de actividad delictiva (BBC News, 2019). Los servicios de IA bien diseñados según los criterios de la estrategia AuroraAI de Finlandia, pueden compartir información y unir servicios en torno al usuario.[22] No obstante, si bien esos usos probablemente no den lugar a que exista una concentración de poder equiparable a la del ejemplo del Crédito Social, siguen planteando una serie de cuestiones que los funcionarios públicos deberían considerar:

- La correlación histórica entre ciertas características y un resultado negativo no es prueba de que exista un vínculo causal que perdurará a lo largo del tiempo. Parapetar estas relaciones en las puntuaciones de los datos puede generar estereotipos, discriminación y la percepción de que estas correlaciones no pueden modificarse a través de políticas públicas eficaces o de la elección personal. Por lo tanto, puede causar la estigmatización de personas con ciertas características.
- La IA puede utilizarse para identificar irregularidades y disciplinar a los ciudadanos, por ejemplo, al identificar fraudes en las prestaciones, y así reducir los costos de los

[22] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/event/semic-webinar-artificial-intelligence-and-public-administrations-09-04-2019-1000-1130-cet.

servicios en el contexto de unas finanzas públicas ajustadas. Existe el riesgo de que, con el uso de algoritmos para estos fines, se despersonalizarán los servicios públicos que antes prestaban los trabajadores sociales y se creará un sistema punitivo que podría afectar negativamente a los más vulnerables de la sociedad. A estas personas también les puede resultar difícil buscar reparación si se les identifica incorrectamente como infractores de las normas (Shafique, 2018; Whittaker et al., 2018)). Por consiguiente, los usuarios, en particular los grupos marginados, podrían mostrar frustración como resultado del "aumento de las cargas administrativas" en la forma de una burocracia confusa y reglamentos complejos que crean barreras para el acceso a los servicios (Herd y Moynihan, 2018).

- En relación con esto, es probable que tenga que haber algunas concesiones entre la prestación de servicios públicos universales que se consideran un derecho de los ciudadanos, por una parte, y los servicios adaptados en función de las características captadas en las puntuaciones de los datos, por otra. Si bien estos últimos pueden ofrecer servicios más específicos y apropiados, dichos servicios pueden dar lugar a una mayor complejidad para los usuarios. Además, si los servicios dejan de ser universales, ello podría tener como consecuencia una reducción del apoyo a los servicios de los que no todos se benefician (Shafique, 2018).

Los gobiernos deberán tener en cuenta tanto el sesgo como la equidad al explorar el potencial de las políticas y los servicios impulsados por la IA.

Garantizar la recopilación, el acceso y el uso éticos de datos de calidad

Como se discutió en el Capítulo 2, los datos son los cimientos de la IA. Una clara estrategia de datos que les permita a los gobiernos acceder a datos ricos, precisos y útiles, mantener la privacidad, y que se ajuste a las normas sociales y éticas será una condición básica necesaria para el despliegue eficaz de la IA (véanse en el Recuadro 4.19 ejemplos de estrategias de datos y en el anexo A un caso de estudio sobre la Estrategia Federal de Datos de los Estados Unidos y el Plan de Acción relacionado). La IA depende del acceso a datos de calidad, sin embargo, la obtención de esos datos es costosa y administrativamente compleja. Por consiguiente, los gobiernos deben tener una supervisión clara de sus activos existentes y un enfoque estratégico para crear conjuntos de datos de calidad en las esferas que están listas para el desarrollo de IA, así como enfoques para obtener datos de fuentes externas, como el sector privado.

Recuadro 4.19: Ejemplos de estrategias nacionales de datos existentes de distintos gobiernos

Varios países han establecido estrategias para aprovechar sus activos de datos. Esto puede implicar la creación de normas congruentes, protocolos de intercambio de datos entre gobiernos y la apertura de datos gubernamentales. Por ejemplo, el Gobierno de Uruguay desarrolló una estrategia de datos que promueve los datos como un activo fundamental para todas las operaciones gubernamentales, y propone un enfoque de sistemas para la recopilación, gestión y administración de los datos. Uruguay también ha puesto en marcha una plataforma de interoperabilidad diseñada para facilitar y

Recuadro 4.19: Ejemplos de estrategias nacionales de datos existentes de distintos gobiernos *(Cont.)*

promover los servicios digitales del gobierno, y mejorar la integración entre distintos organismos del sector público.

El Gobierno de Nueva Zelanda también cuenta con un conjunto de principios para la gestión de datos que incorpora los principios de datos abiertos.

La Estrategia Federal de Datos [*Federal Data Strategy*] de los Estados Unidos, establece principios y prácticas congruentes "para ofrecer un enfoque más coherente de la administración, el uso y el acceso a los datos federales" (véase el caso de estudio del Anexo A).

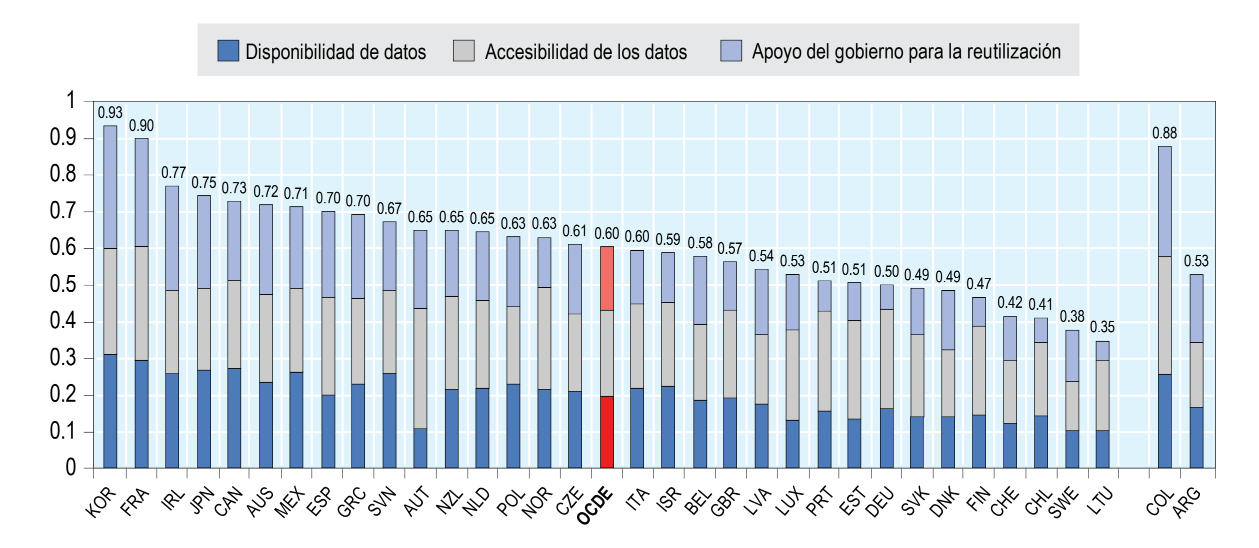
Además, podría haber justificación para la armonización internacional de las normas digitales y de datos. Actualmente existen muchos foros para compartir las mejores prácticas en materia de gobierno digital y datos, entre ellos el Grupo Digital 9 y el Grupo de Trabajo de la OCDE de Líderes de Gobierno Electrónico (E-Leaders).

Fuente: www.oecd-ilibrary.org/governance/a-data-driven-public-sector_09ab162c-en; www.agesic.gub.uy/innovaportal/v/1711/9/agesic/que-es.html?idPadre=3922; www.digital.govt.nz/standards-and-guidance/data-2/data-management; https://strategy.data.gov; www.digital.govt.nz/digital-government/international-partnerships/the-digital-9.

En años recientes se ha observado un creciente interés por la interoperabilidad, la apertura y el intercambio de datos del sector público. En todo el mundo, muchos gobiernos han realizado grandes inversiones en la elaboración de políticas de Datos Abiertos Gubernamentales (DAG) dirigidas a aumentar la apertura, la utilidad y la reutilización de los datos gubernamentales, así como de los portales de datos abiertos relacionados. Debido a que existe el potencial de que los datos sirvan de alimento para la IA en todos los sectores, los gobiernos están ampliando sus políticas e iniciativas en materia de DAG. Como se discutió en el Capítulo 3, la apertura de los datos gubernamentales es uno de los temas más frecuentes en las estrategias nacionales de IA. Las investigaciones recientes de la OCDE indican que la publicación de DAG en formatos legibles por máquina, es una de las más altas prioridades de las estrategias nacionales de los Gobiernos Digitales. Sus esfuerzos se miden mediante el índice OURdata de la OCDE (véase la Figura 4.3).[23] Los gobiernos pueden utilizar los datos internos y DAG de sus propios países, así como los datos disponibles de otros gobiernos.

[23] La OCDE ha publicado varios informes y otros productos relacionados con el estado actual de las estrategias e iniciativas de los DAG, así como los desafíos asociados, las lecciones aprendidas, los casos de estudio y las recomendaciones. Para obtener más información véase www.oecd.org/gov/digital-government/open-government-data.htm

Figura 4.3: Índice de datos abiertos, útiles y reutilizables (OURdata) de la OCDE, 2019



Nota: No se dispone de datos sobre los Estados Unidos, Hungría, Islandia y Turquía. Información sobre los datos de Israel: http://dx.doi.org/10.1787/888932315602.
Fuente: OECD Open Government Data Survey 2018; OCDE (2019), Government at a Glance 2019, OECD Publishing, París, https://doi.org/10.1787/8ccf5c38-en.

Si bien el sector público posee enormes recursos de datos e información, los organismos gubernamentales también pueden aprovechar fuentes externas de información para cumplir sus misiones de forma más eficiente. Por ejemplo, el análisis, con frecuencia en tiempo real, de una amplia gama de datos reunidos o generados por el sector privado puede crear oportunidades para la IA. La industria produce enormes cantidades de datos, frecuentemente en formatos legibles por máquina, diseñados para que éstas puedan consumirlos fácilmente. Cada vez con mayor frecuencia, las empresas publican sus datos como datos abiertos, que los gobiernos y otras entidades pueden consumir sin tener que preocuparse por situaciones de adquisición y gastos. En otros casos, es posible que los gobiernos tengan que comprar el acceso a los datos del sector privado o tratar de ofrecer incentivos a las empresas para que compartan sus datos (OCDE, 2007a).

La OCDE ha venido realizando una amplia labor para mejorar el acceso a los datos y su intercambio (EASD, por sus siglas en inglés), incluido el intercambio de datos de otros sectores con el sector público.[24] En particular, la OCDE actualmente trabaja en la elaboración de principios generales para mejorar el acceso a los datos y su intercambio en toda la economía de manera coherente, mismos que deberían publicarse a finales de 2019 o principios de 2020. Además de los esfuerzos de la OCDE por elaborar principios internacionales, los gobiernos también están poniendo de su parte para aprovechar mejor los datos del sector privado. Por ejemplo, los gobiernos de Francia y Alemania se están centrando en diferentes enfoques para facilitar el paso del intercambio de datos del sector privado al sector público (Recuadro 4.20).

También a nivel de gobierno local se están realizando esfuerzos para obtener datos del sector privado. Por ejemplo, Barcelona (España) ha empezado a añadir cláusulas en los contratos de contrataciones públicas relativas a la soberanía de los datos y que garantizan la propiedad

[24] www.oecd.org/sti/ieconomy/enhanced-data-access.htm.

pública de los datos recogidos de los usuarios, incluso por empresas privadas. En una entrevista realizada en mayo de 2018 con la revista *Wired*, la Comisionada de Tecnología del Ayuntamiento de Barcelona, Francesca Bria, declaró, "tenemos un gran contrato con Vodafone, y cada mes Vodafone tiene que entregar datos legibles por máquina al ayuntamiento. Antes, eso no sucedía. Simplemente tomaban todos los datos y los usaban para su propio beneficio". (Graham, 2018) Actualmente, la ciudad está desarrollando herramientas en colaboración con las demás ciudades como parte del proyecto financiado por la UE, DECODE, el cual pretende crear herramientas que les permitan a las personas controlar si mantienen su información personal privada o la comparten para beneficio del interés público.[25]

Recuadro 4.20: Planes nacionales para acceder a los datos del sector privado

Los planes de Francia para los datos de interés público en poder del sector privado

En marzo de 2018, el presidente Emmanuel Macron presentó la visión y la estrategia del país para hacer de Francia un líder en materia de IA. Como muchas estrategias nacionales, tiene un enfoque significativo en la apertura de los datos gubernamentales. Sin embargo, también examina los planes de apertura de los datos del sector privado. La estrategia está respaldada por un informe escrito por Cédric Villani, matemático y miembro del Parlamento. Entre otras cosas, propone la creación de leyes que abran de forma gradual el acceso a algunos conjuntos de datos, caso por caso y por sector, para beneficio del interés público. Dependiendo de la sensibilidad de los datos, esto podría llevarse a cabo de una de dos maneras: haciendo que los datos sean accesibles sólo para el gobierno, o haciendo que los datos sean accesibles de forma más amplia, por ejemplo, a otros agentes económicos.

Los planes de Alemania para la infraestructura de intercambio de datos

En noviembre de 2018, Alemania publicó su Estrategia de Inteligencia Artificial, denominada "IA Hecha en Alemania" [*AI Made in Germany*]. La estrategia exhorta a que se explore la creación de una infraestructura fiable para el intercambio de datos con la finalidad de ayudar a que el sector privado sienta mayor seguridad con respecto a compartir datos con el gobierno. La creación de esta infraestructura sería una empresa conjunta e incluiría a representantes del sector privado, la ciencia y el sector público al amparo de normas abiertas e interoperables.

Fuente: www.aiforhumanity.fr/en, www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf; www.datainnovation.org/2018/08/countries-can-learn-from-frances-plan-for-public-interest-data-and-ai.

Actualmente, la IA posibilita el ingreso de una variedad más amplia de datos a los algoritmos para informar las políticas públicas, algo que previamente no era posible. La elaboración de políticas basadas en evidencias ha dependido durante mucho tiempo de la recolección y el análisis de información para configurar la elaboración y la aplicación de políticas. Sin embargo, por lo general esta información ha adoptado la forma de datos estructurados, como las encuestas. La IA también permite incorporar datos no estructurados, como por

25 https://decodeproject.eu.

ejemplo imágenes y texto abierto obtenidos a partir de interacciones en redes sociales. También puede aprovechar la información generada a través de la prestación de servicios digitalizados. De esta forma, crea oportunidades para mejorar la definición de los problemas y la formulación de políticas, y permite una comprensión más rápida, profunda y precisa de las preferencias y el contexto de los ciudadanos.

Sin embargo, esos datos también crean nuevos desafíos para los encargados de crear las políticas. Los datos inadecuados generarán sistemas de IA que harán recomendaciones en torno a decisiones equivocadas. Si los datos reflejan las desigualdades presentes en la sociedad, entonces la aplicación de la IA podría reforzarlas, lo cual a su vez podría distorsionar los desafíos y las preferencias en materia de políticas (Pencheva, Esteve y Mikhaylov, 2018). Si se ha entrenado a la IA con datos de un subconjunto de la población que tiene características diferentes a las de la población en general, entonces el algoritmo puede generar resultados sesgados o incompletos. Esto podría llevar a que las herramientas de la IA refuercen las formas de discriminación existentes, como el racismo y el sexismo.[26] También puede ser más difícil anonimizar los datos no estructurados y, de esta forma, se podrían socavar las normas de privacidad.

La Jerarquía de Necesidades de la Ciencia de Datos, establecida en el Capítulo 2, sienta bases claras para comenzar a conceptualizar la recopilación, almacenamiento, transformación, análisis e implementación, los pasos a seguir para tomar datos en bruto y utilizarlos para generar nuevos conocimientos e información. Es probable que cada uno de esos pasos implique cuestiones éticas y jurídicas, así como técnicas.

Como mínimo, el uso que hagan los gobiernos de la IA debe ajustarse a las leyes nacionales de protección de datos. El Reglamento General de Protección de Datos (RGPD), que entró en vigor en mayo de 2018, crea normas uniformes de protección de datos para todos los organismos que operan en la Unión Europea con el fin de dar a las personas un mayor control sobre sus datos personales y crear un "campo con igualdad de oportunidades" para las empresas.[27] Entre otras disposiciones, establece normas en torno a la creación de capacidades de protección de datos en las distintas dependencias, promueve la transparencia y da a los ciudadanos la posibilidad de decidir sobre las formas de almacenamiento de sus datos personales y los usos que se les puede dar a dichos datos.

Por ejemplo, en el Artículo 5 del RGPD se describe el concepto de "limitación de la finalidad", que restringe los términos en los que las distintas dependencias pueden reutilizar los datos que han recolectado o adquirido en otro lugar. En particular:

> *Si su empresa/dependencia ha recopilado los datos tras haber solicitado el consentimiento para hacerlo o en cumplimiento con un requisito legal, no será posible llevar a cabo ningún otro procesamiento más allá de lo que abarca el consentimiento original o las disposiciones de la ley. Si se desea llevar a cabo un procesamiento adicional, sería necesario obtener un nuevo consentimiento o un nuevo fundamento jurídico.*[28]

[26] www.digital.nsw.gov.au/digital-transformation/policy-lab/artificial-intelligence.
[27] https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en.
[28] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr/purpose-data-processing/can-we-use-data-another-purpose_en.

Los gobiernos deberán tener en cuenta estas leyes durante el diseño y la elaboración de estrategias de datos e iniciativas de IA. Si bien este enfoque puede frenar el despliegue de la IA en el futuro inmediato, podría sentar las bases para una IA más ética e inclusiva a largo plazo. Algunos países ajenos a la Unión Europea han adoptado el RGPD de forma voluntaria, lo cual significa que éste puede llegar a constituir la base de una norma mundial para la de protección de datos. Esto se ajusta al énfasis que la Comisión Europea hace sobre el desarrollo de la IA de una manera que fomente la confianza del público mediante el establecimiento de normas y protecciones sólidas, como se ha expuesto anteriormente.

Las normas culturales influirán en las opiniones populares con respecto a la privacidad, los datos que es ético utilizar y las restricciones o permisos que se deben exigir. El equilibrio entre la privacidad y el uso de los datos para mejorar los servicios y hacerlos más personalizados será difícil de capturar y variará según el contexto. Generar un consenso estable en toda la sociedad sobre las distintas concesiones, por ejemplo, entre privacidad, transparencia y calidad del servicio (Janssenand van den Hoven, 2015), será un desafío. Cuando la confianza en el gobierno es escasa, es probable que haya oposición a la recolección de grandes cantidades de datos y a su utilización en formas que no son claras para el público. Por ejemplo, la tecnología de reconocimiento facial es un método de recopilación de datos que ha puesto de relieve las preocupaciones en torno al equilibrio correcto entre servicios más eficaces y la privacidad y el posible sesgo (véase el Recuadro 4.21).

Recuadro 4.21: La tecnología de reconocimiento facial, la privacidad y las inquietudes en torno a los prejuicios

La tecnología de reconocimiento facial puede tener muchos usos transformadores. Por ejemplo, su uso para pagar los viajes en metro está en fase de pruebas en Futian, China. Sin embargo, la tecnología se ha convertido en un imán para las preocupaciones sobre la privacidad. A medida que esta tecnología ha avanzado, se ha vuelto cada vez más capaz de identificar rostros en una multitud a través del uso de datos de imagen facial. Por ejemplo, al cotejar las imágenes de las cámaras de vigilancia con las bases de datos de la policía, puede ofrecer vigilancia en tiempo real y mejorar la seguridad identificando a sospechosos de delitos o a personas desaparecidas, entre otras aplicaciones. En China, el reconocimiento facial se ha complementado con el "análisis de la marcha", que identifica a las personas por su forma de caminar. Sin embargo, a los defensores de la privacidad les preocupa que esto permita a los gobiernos recopilar una enorme cantidad de información sobre los ciudadanos sin su consentimiento, la cual podría utilizarse con distintos fines.

Además, la tecnología de reconocimiento facial que se entrena con conjuntos de datos que no son suficientemente diversos puede reducir la precisión de la identificación para algunos grupos, lo que aumenta el riesgo de falsos positivos. Por ejemplo, las fuerzas policiales del Reino Unido han recibido críticas por no haber comprobado el impacto del origen étnico en la precisión de las predicciones. En un estudio del MIT, cuyos resultados se impugnan, se determinó que múltiples herramientas de reconocimiento facial son menos precisas para las personas de raza negra y las mujeres, lo que da lugar a posibles sesgos por motivos de género y origen étnico.

Recuadro 4.21: La tecnología de reconocimiento facial, la privacidad y las inquietudes en torno a los prejuicios *(Cont.)*

También hay casos en los que procedimientos inadecuados seguidos por las fuerzas policiales dieron lugar a la utilización de datos de entrada de mala calidad, lo que melló considerablemente la precisión del software de reconocimiento facial. Por ejemplo, las fuerzas policiales de los Estados Unidos han intentado cotejar retratos hablados de los sospechosos, fotogramas de CCTV de mala calidad, imágenes mejoradas por computadora e incluso una foto de una celebridad que tenía un gran parecido con un sospechoso con bases de datos de imágenes. Estos ejemplos sugieren que se requieren reglas más claras sobre la forma exacta en que debe utilizarse el software y para aclarar si una coincidencia es justificación suficiente para la detención.

En un contexto en el que la tecnología cambia rápidamente y con bajos niveles de confianza en el gobierno, existe la preocupación de que esta tecnología otorgue demasiado poder al sector público. Es probable que casos controvertidos como éstos precipiten un debate social sobre si el uso de la tecnología de reconocimiento facial es compatible con el respeto de la autonomía individual y, en caso de ser así, qué salvaguardas deben establecerse para proteger los valores liberales. Es probable que los ciudadanos exijan que se les consulte de forma adecuada sobre si la tecnología se está utilizando de una manera que pudiera afectarlos. Como respuesta contra su uso, San Francisco se ha convertido en la primera ciudad de los Estados Unidos en prohibir el uso del reconocimiento facial por parte del ayuntamiento.

Fuente: https://towardsdatascience.com/how-ethical-is-facial-recognition-technology-8104db2cb81b; www.scmp.com/tech/innovation/article/3001306/you-can-soon-pay-your-subway-ride-scanning-your-face-china; www.bbc.co.uk/news/technology-47117299; https://medium.com/@AINowInstitute/after-a-year-of-tech-scandals-our-10-recommendations-for-ai-95b3b2c5e5; www.flawedfacedata.com; www.americaunderwatch.com; www.bbc.co.uk/news/technology-48222017; www.vox.com/future-perfect/2019/5/16/18625137/ai-facial-recognition-ban-san-francisco-surveillance; www.sfchronicle.com/politics/article/SF-could-ban-facial-recognition-software-13842657.php.

Garantizar que el gobierno tenga acceso a los fondos, a la capacidad interna y externa y a la infraestructura

Los mecanismos de financiamiento son un factor importante de los usos de la IA en el sector público. Incluso las iniciativas más sencillas necesitan tener acceso a algún nivel de financiamiento y apoyo financiero para materializarse. La disponibilidad y la naturaleza de este financiamiento pueden contribuir en gran medida al éxito final de las innovaciones basadas en la IA (OCDE, 2007b).

El acceso a la capacidad también es fundamental. Las diferencias en los niveles iniciales de madurez y capacidad institucionales en el gobierno, el mundo académico, la sociedad civil y el sector privado, darán lugar a limitaciones en el acceso del gobierno a profesionistas de primera categoría y requerirán diferentes enfoques estratégicos para recoger los frutos de la IA. Por ejemplo, es probable que el despliegue efectivo de la IA dependa de aptitudes y capacidades técnicas y de transformación de los servicios que probablemente no existan actualmente en el sector público. Podría ser complicado formar esas aptitudes internamente, pero también puede resultar difícil obtenerlas de manera externa debido a los engorrosos

procesos de contratación y adquisición con el sector privado o a los inadecuados mecanismos de colaboración con los círculos académicos y la sociedad civil.

Como se mencionó anteriormente en este capítulo, la estrategia adoptada por los gobiernos para desarrollar a los profesionistas de primera categoría de la IA debería centrarse no sólo en las aptitudes técnicas sino también en el desarrollo multidisciplinario de capacidades relativo a las implicaciones sociales, éticas y jurídicas de la IA, y en cambiar la postura y las formas de trabajar necesarias para colaborar con los equipos mixtos y la IA (AI Now, 2018).[29]

Los gobiernos deberían tratar de abordar esos retos técnicos y no técnicos a través de enfoques innovadores en materia de capacitación, contratación, adquisición y asociación. Para alcanzar el potencial de la IA para el sector público, los gobiernos tendrán que aplicar una combinación de enfoques. Sin embargo, independientemente de esta combinación, al mismo tiempo siempre tendrán que mantener un papel único e integral en la fijación de rumbos, la creación de normas y la vigilancia del cumplimiento de las políticas y leyes, ya que siempre será responsabilidad del gobierno garantizar el diseño y el uso adecuados de la IA en el sector público. Por último, los gobiernos deberán valorar sus necesidades actuales en materia de infraestructura técnica en relación con sus ambiciones y asegurarse de que se dispone de una infraestructura moderna para poder avanzar en la exploración de la IA.

Asegurar la disponibilidad de recursos de financiamiento

El financiamiento es un factor importante para que las iniciativas y estrategias de inteligencia artificial del sector público tengan éxito. Como se mencionó en el Capítulo 1, la reducción tanto del interés como del financiamiento sustentable ha sido un factor importante en los previos inviernos de la IA. Para contribuir a garantizar que el sector público pueda recoger los frutos de la IA, los gobiernos tendrán que establecer planes de financiamiento y asegurar la disponibilidad de recursos en los planes fiscales. El Fondo de Modernización Tecnológica (TMF) de los Estados Unidos, el cual se analizó en secciones anteriores de este capítulo, y la inversión de EUR 100 millones de Finlandia para el periodo 2020-2023 destinada a poner en marcha de 10 a 20 servicios gubernamentales basados en la IA (véase el estudio de caso sobre la Estrategia Nacional de IA de Finlandia en el Anexo A), son ejemplos de ello. El documento de la OCDE sobre *Estado de la Técnica de las Tecnologías Emergentes en el Sector Público* (Ubaldi et al., 2019) presenta un panorama más detallado sobre el financiamiento gubernamental de la IA para el sector público.

Crear la capacidad interna

Es probable que la transformación extensa de la IA tenga repercusiones sustanciales en las aptitudes necesarias para prestar eficazmente los servicios públicos. Estos cambios incluirán lo siguiente:

- Los altos dirigentes deberán saber cómo maximizar el valor de la IA en los servicios públicos, incluso mediante los cambios organizacionales necesarios para obtener el máximo beneficio de las tecnologías basadas en la IA.
- Los directivos tendrán que comunicar a los empleados los beneficios y desafíos relacionados con la IA.

[29] www.pwc.com/us/en/services/consulting/library/artificial-intelligence-predictions/employer-impact.html.

- Los administradores de los servicios supervisarán su prestación mediante una puesta en marcha efectiva.
- Es posible que sea necesario contar con conocimientos técnicos internos para que el gobierno pueda ser un líder intelectual en este ámbito y negociar de forma eficaz con los contratistas.
- El personal deberá tener las habilidades y capacidades necesarias para trabajar en conjunto con la IA, así como interpretarla y complementarla.

En todos estos niveles, el desarrollo de una fuerza de trabajo de IA diversa que refleje la composición social de la población, mediante prácticas de contratación y la creación de una cultura inclusiva, será una salvaguarda crucial contra las prácticas poco éticas, los prejuicios y el pensamiento de grupo (du Preez, 2018). Las inversiones en capacitación dirigida a mejorar los niveles de alfabetización en materia de IA en toda la dependencia pueden reducir la ansiedad en el lugar de trabajo con respecto a lo que su uso representa para el personal. La Academia Digital [*Digital Academy*] de la Escuela de la Función Pública de Canadá [*Canada's School of Public Service*] (véase el Recuadro 4.22), ofrece un ejemplo de los distintos niveles de capacitación en materia de IA y tecnologías digitales diseñados para adaptarse a las necesidades de los diversos grupos de funcionarios.

Recuadro 4.22: Academia Digital de la Escuela de la Función Pública de Canadá

La Academia Digital es un organismo de enseñanza de la Escuela de la Función Pública de Canadá. Ésta ofrece recursos a los funcionarios públicos para mejorar sus operaciones a través de la prestación de servicios digitales. El programa educativo de la Academia forma parte de un plan más amplio para la reforma del sector público con miras a la creación de un servicio público ágil, inclusivo y equipado.

La Academia Digital ofrece formación para funcionarios de todos los niveles jerárquicos y con diferentes niveles de conocimientos especializados. Utiliza desafíos y problemas de la vida real y una combinación de eventos, aprendizaje en línea y podcasts (denominados busrides.ca y que están diseñados para dar presentaciones rápidas sobre temas relacionados con los servicios digitales que ofrece el gobierno). Las oportunidades de aprendizaje siguen tres niveles:

1. El nivel de **Fundamentos Digitales [*Digital Foundations*]** está dirigido a todos los funcionarios públicos sin importar su nivel de especialización. Su objetivo es brindar información oportuna sobre el mundo digital que afectará a la forma en que hacen su trabajo e incluso cómo viven su vida.
2. El nivel de **Digital Premium**, o los materiales especializados para profesionales, se centra en los datos, el diseño, el desarrollo, la IA y el Aprendizaje Automático, las DevOps, también conocidas como operaciones de desarrollo, y la tecnología disruptiva.
3. El nivel de **Liderazgo Digital** [***Digital Leadership***] tiene por objeto desarrollar las aptitudes y la mentalidad digitales de quienes dirigen el diseño y la prestación de servicios, por no mencionar el cambio de cultura necesario para "hacer lo digital" con éxito en la administración pública federal.

Fuente: www.csps-efpc.gc.ca/About_us/Business_lines/digitalacademy-eng.aspx.

Se recomienda a los gobiernos estudiar las formas de aumentar los conocimientos especializados de los funcionarios públicos en todos los niveles, centrándose en la creación de un cuadro de dirigentes de alto nivel con conocimientos tecnológicos que pueda abogar por el despliegue de la IA en el gobierno. Los dirigentes de alto nivel, incluso a nivel político, deberán tener una comprensión estratégica de lo que puede hacer la IA y saber cómo identificar los tipos de problemas que se pueden abordar con ésta, así como las preguntas clave que deben formularse para garantizar una supervisión eficaz de los resultados (Agrawal, Gans y Goldfarb, 2018). Es posible que esto no requiera conocimientos técnicos profundos, pero sí la capacidad de actuar como un interlocutor o traductor capaz de comprender los aspectos técnicos, éticos y jurídicos relativos a la puesta en marcha de una IA viable y ética, y de combinarlos con su entendimiento del funcionamiento de los servicios y las administraciones públicas (du Preez, 2018).[30]

Será necesario que los administradores de los servicios habilitados por IA tengan conocimientos técnicos más profundos, incluso si son contratistas externos quienes prestan los servicios. El conocimiento de la IA, habilidades de negociación eficaces y la experiencia en el sector ayudarán a los administradores de servicios a diseñar contratos adecuados que garanticen una supervisión eficaz y a exigir responsabilidad por parte de los contratistas externos. Estas aptitudes también pueden ayudar a que las soluciones de IA propuestas se evalúen de forma correcta con la finalidad de verificar su idoneidad para el propósito y el precio exacto. Los administradores de los servicios tendrán que trabajar en estrecha colaboración a través de redes de representantes del sector privado, la sociedad civil y el mundo académico, aprovechando sus conocimientos y colaborando eficazmente. Sin embargo, también deberán evitar la influencia indebida de interesados externos, además de no desviar su atención de los objetivos de su organismo ni de la maximización del valor público. Por lo tanto, el obtener conocimientos técnicos internos y de capacidad de negociación de manera constante ayudará a los gobiernos a poner en marcha servicios y a colaborar de forma eficaz con los interesados externos.[31]

Es crucial que tanto los dirigentes como los administradores de alto nivel estén listos para manejar el cambio. Esta preparación incluirá la comunicación con los empleados sobre los desafíos y beneficios asociados con la IA, ya que se ha demostrado en investigaciones que esto reduce los niveles de estrés de los empleados a medida que sus trabajos y tareas cambian como resultado de la automatización y la IA (Eggers, Schatsky y Viechnicki, 2017; Viechnicki y Eggers, 2017). También se deberá incluir la aplicación de los cambios organizacionales, estructurales y procedimentales necesarios para complementar la adopción de la IA. Asimismo, deberán ser capaces de evaluar la IA para determinar si las aplicaciones en cuestión funcionan de manera adecuada y para mitigar los posibles riesgos y otras consecuencias adversas.

Podría haber casos en los que los gobiernos deseen desarrollar conocimientos técnicos internos para ayudarles a asumir un papel de liderazgo en el espacio de la IA. Hay razones de peso para desarrollar el talento de IA dentro del gobierno, especialmente en áreas que pueden conllevar a consecuencias graves en materia de seguridad o cuando haya

30 www.pwc.com/us/en/services/consulting/library/artificial-intelligence-predictions/functional-specialists.html.

31 https://joinup.ec.europa.eu/sites/default/files/event/attachment/2019-04/SEMIC_Webinar%20on%20AI%20and%20PA_Presentation_AI%20projects%20in%20PA%20in%20Japan_Kenji%20Hiramoto%20%28JP%29_09-04-2019.pdf.

oportunidades particulares para compartir el aprendizaje en toda la dependencia. Los reglamentos de protección de datos reflejan que con frecuencia es más fácil trasladar a las personas que los datos. En estos casos, las metodologías podrían incluir la incorporación de contratistas externos en el sector público, la exploración de comisiones internas o la creación de capacidad interna (Mikhaylov, Esteve y Campion, 2018).

Un método para fortalecer la capacidad interna consiste en crear talento interno en materia de IA mediante la especialización profesional del personal existente, como los estadísticos y los científicos de datos, con las habilidades y aptitudes pertinentes, y explorando la posibilidad de realizar comisiones externas a organizaciones innovadoras para crear conocimientos para toda la dependencia. En el Recuadro 2.5 del Capítulo 2, que trata sobre la Automatización Robótica de Procesos en los Estados Unidos, se ofrece un ejemplo de reinversión de los ahorros en los costos derivados de la automatización en la especialización profesional del personal existente, para que éste se pueda desempeñar en funciones más estratégicas.

Otro método es reclutar expertos para el gobierno. Esto a menudo puede resultar difícil, ya que los gobiernos suelen tener normas estrictas sobre la contratación en el sector público. Los conocimientos en materia de IA tienen una gran demanda y el sector público podría encontrar dificultades para contratar y retener al personal si no le es posible ser flexible y competir con los salarios del sector privado. Los gobiernos podrían aprovechar la experiencia (o incluso autoridades especiales de contratación) de sus iniciativas anteriores para conseguir especialistas, como científicos, administradores de proyectos experimentados y economistas. Las opciones incluyen intentar atraer al personal con la oferta de incentivos, condiciones de trabajo flexibles, oportunidades de desarrollo únicas y experiencias que ayuden a su desarrollo profesional a largo plazo. Algunos gobiernos también han aprovechado el importante impacto cívico del servicio público como herramienta de reclutamiento. Por ejemplo, el Servicio Digital de los EE. UU. [*US Digital Service*][32] recluta especialistas en tecnología para el gobierno para fortalecer las misiones sociales a través de "giras de servicio cívico" de duración limitada. Los mecanismos de contratación formalizados con programas de desarrollo centralizados también pueden ayudar a proporcionar conocimientos básicos sobre temas importantes y a desempeñar funciones en todo el gobierno (véase el ejemplo del Recuadro 4.23).

Recuadro 4.23: El programa Fast Stream del Reino Unido

Fast Stream es uno de los programas de desarrollo para graduados más grandes de la OCDE en esta región. En 2015, 21,135 candidatos compitieron por 967 nombramientos en 12 corrientes diferentes de especialistas y generalistas. Una vez seleccionados, el programa dota a los participantes de los conocimientos, habilidades y experiencia que necesitan para convertirse en los futuros dirigentes de la administración pública. El desarrollo personal de los participantes seleccionados de Fast Stream se logra mediante un programa de puestos cuidadosamente gestionados y contrastados, complementados con un aprendizaje formal y otras formas de apoyo como el entrenamiento, la tutoría y el aprendizaje práctico. El programa Fast Stream del Reino

[32] https://usds.gov.

Recuadro 4.23: El programa Fast Stream del Reino Unido *(Cont.)*

Unido, promueve la diversidad y la inclusión y genera un informe anual con datos y análisis que demuestran la variedad de solicitantes. El informe constituye un buen ejemplo de análisis de recursos humanos basado en datos. En 2015 se elaboró un informe adicional diseñado para comprender los factores que explicaran el patrón socioeconómico de los candidatos al programa Fast Stream, y examinar las razones por las que los candidatos procedentes de entornos socioeconómicos más bajos tienen menos probabilidades de presentar una solicitud y de tener éxito. Esta investigación saca a la luz información pertinente para la administración pública en general, y evidencia para aprovechar y hacer recomendaciones con miras a mejorar la diversidad socioeconómica.

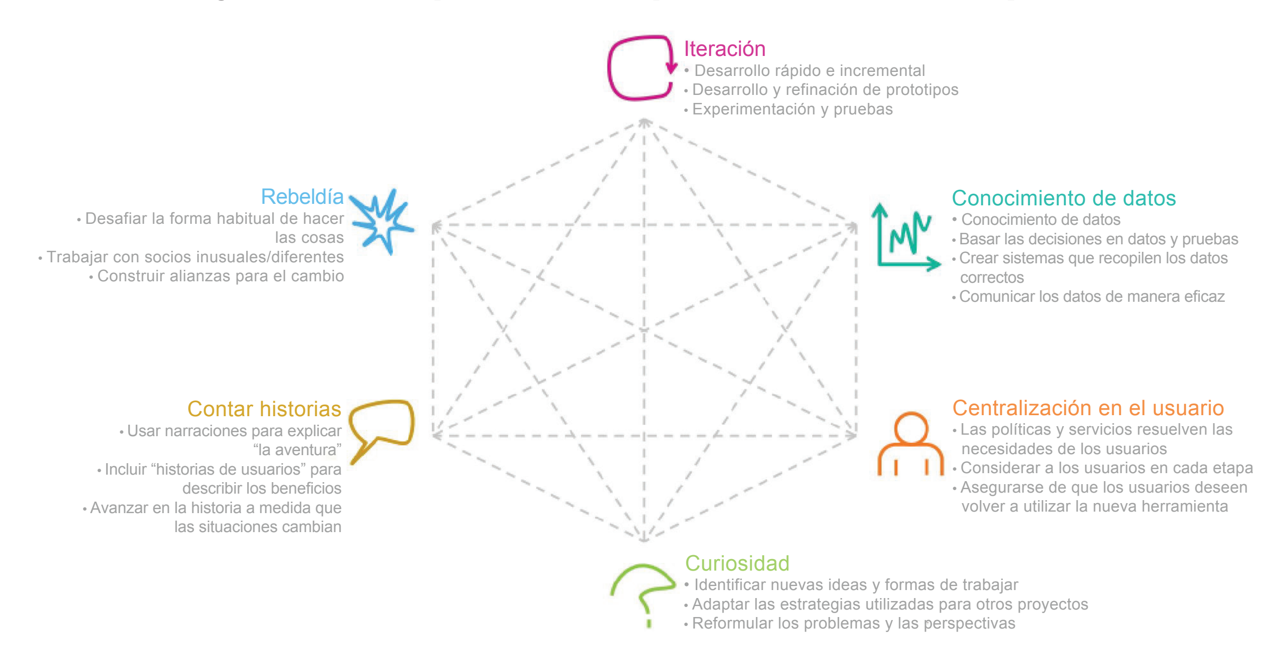
Fuente: The Bridge Group (2016), *Socio-Economic Diversity in the Fast*Stream, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/497341/BG_REPORT_FINAL_PUBLISH_TO_RM__1_.pdf (consultado el 26 de septiembre de 2019).

Además de los conocimientos técnicos (p. ej., ciencia de datos, desarrollo de código, etc.) y la capacidad de interpretar los resultados de los algoritmos, la aparición de la IA aumenta el valor de las competencias complementarias.[33] Estas son las habilidades necesarias para realizar tareas que la IA no puede hacer, como la capacidad de hacer juicios que tomen en cuenta una variedad objetivos basados en datos escasos, la creatividad y la inteligencia emocional. El OPSI desarrolló las *Competencias básicas para la innovación del sector público* [*Core Skills for Public Sector Innovation*] con el fin de brindar orientación a los gobiernos en la creación de competencias para el siglo XXI. No todos los funcionarios públicos necesitarán hacer uso o aplicar estas competencias en su trabajo del día a día. Sin embargo, para ofrecer un servicio público moderno, todos los funcionarios deben contar con al menos cierto nivel de conocimiento sobre estas seis áreas a fin de impulsar aún más la innovación. Éstas son:

- **Iteración** - desarrollo incremental y experimental de políticas, productos y servicios
- **Conocimientos de datos** - garantizar que las decisiones se basen en los datos y que éstos no sean sólo un concepto secundario
- **Centralización en el usuario** - los servicios públicos deben centrarse en resolver y atender las necesidades de los usuarios
- **Curiosidad** - buscar y probar nuevas ideas o formas de trabajar
- **Contar historias** - explicar el cambio de una manera que sea posible obtener apoyo
- **Rebeldía** - desafiar el statu quo y trabajar con socios inusuales.

[33] www.pwc.com/us/en/services/consulting/library/artificial-intelligence-predictions/employer-impact.html.

Figura 4.4: Seis competencias básicas para la innovación del sector público



Fuente: https://oe.cd/innovationskills.

Además de estas competencias para la innovación, la creatividad, la resistencia personal y las habilidades emocionales seguirán cobrando importancia, especialmente a medida que la naturaleza del trabajo cambie como consecuencia del aumento en la automatización y la adopción de la IA.[34] A medida que la IA se vuelve más barata y más prevaleciente, la innovación y las habilidades emocionales podrían cobrar más importancia en el gobierno. Por ejemplo, el personal que desempeña funciones operativas puede ver cómo su trabajo evoluciona y cambia a medida que la IA se encarga de cada vez más tareas rutinarias, lo que les libera más tiempo para poder centrarse en los casos más complejos y en las relaciones con los ciudadanos. Dependiendo de las decisiones tomadas por los altos dirigentes, este indicador podría dar lugar a servicios más personalizados y de alta calidad o a una reducción del número de puestos que implican tareas que son más sustituibles para la IA. Además de cambiar la naturaleza de muchos trabajos, la IA también es relevante en el sentido de que está cambiando la naturaleza de los mecanismos de los recursos humanos. La IA, por ejemplo, puede utilizarse para vigilar la productividad del personal y fundamentar las decisiones de contratación y gestión del desempeño (Recuadro 4.24).

[34] Para obtener más información sobre las habilidades sociales y emocionales, véase www.oecd.org/education/ceri/study-on-social-and-emotional-skills-the-study.htm

Recuadro 4.24: La IA y la administración de los recursos humanos

La IA tiene el potencial de cambiar significativamente la forma en que las dependencias administran sus recursos humanos, al hacerlas más sensibles y orientadas a los datos. Por ejemplo, la IA podría influir en las decisiones de contratación a través de predicciones sobre los mejores candidatos en función de las características de aquellos que se hayan aprobado anteriormente. Sin embargo, la experiencia de Amazon, en la que un algoritmo de contratación estaba sesgado en contra de las mujeres solicitantes, destaca la importancia que tienen la supervisión rigurosa, los exámenes y una gestión eficaces para identificar los resultados indeseados.

También podría posibilitar una retroalimentación más rápida basada en una gama de datos más amplia que la de los sistemas convencionales de gestión del desempeño, y contrarrestar los sesgos conscientes e inconscientes de los directivos. Aun así, no podemos olvidar que hay riesgos que deben ser mitigados. Tomemos como ejemplo un sistema de IA que se usó en Houston (Texas) para hacer recomendaciones sobre los profesores que deberían recibir un ascenso o ser despedidos con base en los resultados de los exámenes de sus estudiantes, el cual pone de relieve la importancia de que las organizaciones entiendan bien los sistemas y de que las decisiones sean explicables. A menos que se encuentre una forma de resolver o mitigar estas cuestiones, las dependencias pueden ver una caída en la moral de los empleados e incluso podrían verse vulnerables a la impugnación en juicio.

La ansiedad en el lugar de trabajo con respecto al impacto de la IA puede atenderse mediante una descripción de lo que la IA significa para el personal, en la que se aclare la forma en que su situación cambiará dentro de la dependencia en respuesta a la transformación de los servicios posibilitada por la IA.

Fuente: www.forbes.com/sites/insights-intelai/2018/11/29/how-ai-can-help-redesign-the-employee-experience/#19f64c044b34; www.forbes.com/sites/bernardmarr/2017/01/17/the-future-ofperformance-management-how-ai-and-big-data-combat-workplace-bias/#1517089a4a0d; https://ainowinstitute.org/AI_Now_2018_Report.pdf; www.bbc.co.uk/news/technology-45809919.

Los gobiernos deben tener presente que la IA seguirá cambiando la dinámica de trabajo y los requisitos para tener éxito en el sector público en el futuro próximo. Esto enfatiza la necesidad de un aprendizaje y crecimiento continuos. Los gobiernos tendrán que elaborar programas de aprendizaje continuo, y deberán repetir y adaptarlos con el paso del tiempo. El Gobierno de Canadá desarrolló un interesante programa enfocado en las Aptitudes para el Futuro [*Future Skills*] que, aunque originalmente está dirigido a la población canadiense en general, el OPSI considera que un método similar podría también centrarse específicamente en el aprendizaje continuo en el sector público (Recuadro 4.25).

Recuadro 4.25: Aptitudes para el Futuro: Colaborar con los gobiernos y otros interesados en la creación de un ecosistema de desarrollo de aptitudes (Canadá)

El programa Aptitudes para el Futuro forma parte del plan del Gobierno de Canadá para crear una fuerza de trabajo tenaz y con confianza en sí misma que refleje la rápida evolución de la naturaleza del trabajo frente a las tecnologías emergentes. Éste abarca los principios de diseño centrado en el usuario para informar la adopción de prácticas de eficacia comprobada y evidencias sobre los métodos de desarrollo de competencias, a fin de asegurar que las políticas y programas de Canadá estén listas para satisfacer las necesidades cambiantes de los canadienses. Se diseñó en colaboración con los gobiernos provinciales y locales, y recibió aportaciones de una amplia gama de interesados.

En 2017, el Gobierno reconoció la necesidad de "nuevas estrategias para abordar las deficiencias en materia de competencias y apoyar el aprendizaje continuo durante la vida laboral de los canadienses", y se comprometió a invertir en la innovación de las aptitudes de trabajo para fomentar el desarrollo y la medición de las mismas en ese país. El programa resultante de "Aptitudes para el Futuro" prevé una inversión oficial de cuatro años por un total de CAD 225 millones (equivalente a EUR 154 millones), y CAD 75 millones (EUR 67.5 millones) por cada año sucesivo para:

- examinar las principales tendencias que tendrán repercusiones en las economías nacionales y regionales, así como en los trabajadores
- identificar las nuevas competencias que están en demanda actualmente y que lo estarán en el futuro
- desarrollar, probar y evaluar nuevos enfoques para el desarrollo de competencias
- compartir los resultados entre los sectores público, privado y sin fines de lucro para promover un acceso general a las prácticas de eficacia comprobada en todo Canadá.

Como respaldo al programa de Aptitudes para el Futuro, en febrero de 2019, el gobierno puso en marcha dos iniciativas relacionadas:

- *Consejo de Aptitudes para el Futuro* [*Future Skills Council*]: el Consejo está integrado por 15 expertos técnicos y temáticos de todo Canadá, y tiene por objeto garantizar que se tomen en cuenta las necesidades en materia de competencias de todos los habitantes al momento de elaborar las políticas y los programas. La composición es equilibrada en cuanto al género y representa la diversidad social y geográfica de Canadá.
- *Centro de Aptitudes para el Futuro* [*Future Skills Centre*]: Este centro de investigación funciona a distancia del gobierno, y se dedica a financiar proyectos en todo Canadá que desarrollan, someten a pruebas y miden nuevos métodos para la evaluación y el desarrollo de competencias, a fin de acumular evidencias sobre lo que funciona, para quién funciona y en qué condiciones. La mitad del financiamiento del Centro está dirigido exclusivamente a los grupos desfavorecidos y que normalmente no están representados en este ámbito, y hasta el 20% del financiamiento está destinado a atender las necesidades de los jóvenes.

Fuente: https://oecd-opsi.org/innovations/future-skills-engaging-governments-and-stakeholders-to-build-a-skills-development-ecosystem; www.canada.ca/en/employment-social-development/programs/future-skills.html.

Aprovechar los conocimientos especializados externos a través de asociaciones y de la colaboración

Además de crear capacidad interna, los gobiernos pueden recurrir a una red de representantes del sector privado, del mundo académico y de la sociedad civil para aprovechar su experiencia y sus recursos, así como promover el intercambio de conocimientos, a fin de mejorar la toma de decisiones. De hecho, actualmente existen numerosos ejemplos de iniciativas intersectoriales que trabajan para combinar sus capacidades con miras a ofrecer soluciones de IA, como las Oficinas de Análisis de Datos. Éstas suelen constituir un centro de coordinación institucional para la colaboración entre el gobierno local y nacional, las universidades, las empresas de tecnología y las ONG para combinar los datos y abordar los problemas sociales. Por ejemplo, la Oficina del Alcalde de Nueva York para el Análisis de Datos [*Mayor's Office for Data Analytics*] (MODA, por sus siglas en inglés)[35] aprovecha activamente la experiencia de la Universidad de Columbia y la Universidad de Nueva York para desarrollar normas y protocolos de datos (Mikhaylov, Campion y Esteve, 2018).

En el Reino Unido, el Instituto Alan Turing fue creado en 2015 por un consejo de investigación y un grupo de destacadas universidades como el instituto nacional de ciencias para la información e inteligencia artificial. Su Programa de Políticas Públicas permite a las dependencias gubernamentales aprovechar una gran cantidad de conocimientos especializados externos para nutrir de información los servicios públicos y la administración (véase el caso de estudio del Anexo A). El Instituto aprovecha su reputación para atraer representantes académicos, apela al deseo de contribuir al bien común y también ofrece formas flexibles de contribuir que pueden lograrse en torno a otros compromisos profesionales. El Programa de Políticas Públicas, en particular, ha sido considerado un modelo de gran éxito en las entrevistas del OPSI con los interesados en la IA en varios países.[36]

A nivel estratégico, la colaboración intersectorial puede ayudar a los gobiernos a comprender las capacidades existentes y las prioridades de la industria, y a elaborar mejores políticas. La creación de instituciones que faciliten el diálogo puede ayudar a establecer una confianza mutua. Por ejemplo, en Canadá y el Reino Unido se crearon comités de consulta sobre la IA para posibilitar una estrecha colaboración entre el gobierno, el sector privado y el mundo académico (Gobierno de Canadá, 2019b; Gobierno del Reino Unido, 2019a; véase el Recuadro 4.26).

Recuadro 4.26: El Consejo de Inteligencia Artificial [*AI Council*] y el Centro de Ética e Innovación de Datos del Gobierno del Reino Unido

El Gobierno del Reino Unido creó un Consejo Superior de IA que funciona como un comité de expertos independientes para brindar asesoría sobre la forma de estimular la adopción de la IA, promover su uso ético y maximizar su contribución al crecimiento económico.

El Consejo está formado por líderes del sector empresarial, del mundo académico y de la sociedad civil. Se prevé que se volverá el centro de atención para la colaboración intersectorial dentro de la comunidad de la IA con la finalidad de aportar soluciones

[35] www.nyc.gov/analytics.
[36] www.turing.ac.uk/research/research-programmes/public-policy.

Recuadro 4.26: El Consejo de Inteligencia Artificial [*AI Council*] y el Centro de Ética e Innovación de Datos del Gobierno del Reino Unido *(Cont.)*

a las prioridades compartidas, como los datos y la ética, la adopción, las aptitudes y la diversidad.

El Consejo de Inteligencia Artificial trabajará junto con el Centro de Ética e Innovación de Datos, un órgano asesor independiente que analizará y anticipará las oportunidades y los riesgos que plantea la tecnología basada en datos, y presentará sugerencias prácticas y basadas en evidencia para hacerles frente. Esto incluirá revisiones dirigidas a identificar y articular las mejores prácticas para el uso responsable de la tecnología basada en datos, dentro de sectores específicos o para aplicaciones específicas de la tecnología.

Fuente: www.gov.uk/government/news/leading-experts-appointed-to-ai-council-to-supercharge-the-uks-artificial-intelligence-sector; www.gov.uk/government/groups/centre-for-data-ethics-and-innovation-cdei.

A nivel operativo, la prestación de servicios basados en la IA a través de una red de organismos académicos, del sector privado, de la sociedad civil y del sector público puede ayudar al gobierno a aprovechar los conocimientos especializados externos que le permitan mejorar la eficacia y la eficiencia de los servicios públicos. Por ejemplo, en el Reino Unido, el Consejo del Condado de Essex y la Universidad de Essex se asociaron para mejorar los servicios públicos (véase el Recuadro 4.27).

Recuadro 4.27: El Consejo del Condado de Essex y el Instituto de Análisis y Ciencia de Datos de la Universidad de Essex (IADS, por sus siglas en inglés)

El IADS es un ejemplo de la creación de un vehículo institucional para la colaboración intersectorial a nivel de los gobiernos locales. Esta asociación se ve facilitada por el nombramiento conjunto entre las dos organizaciones de un Asesor Científico Principal del Consejo que es también profesor de Políticas Públicas y Ciencia de Datos en la Universidad.

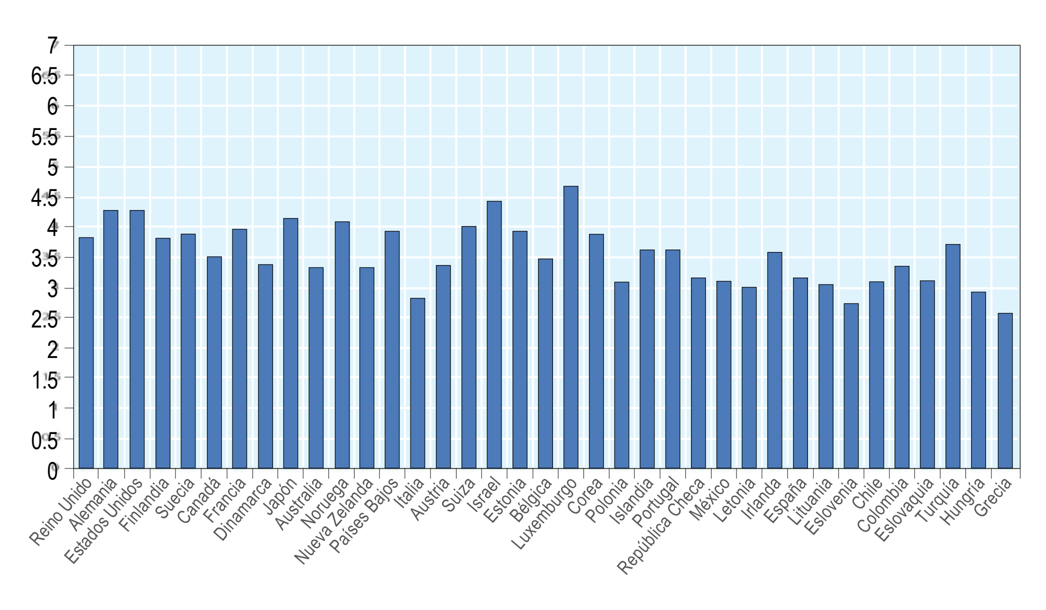
La colaboración permite combinar los datos del sector público y la experiencia con inteligencia artificial de universidades y empresas en beneficio de la comunidad de Essex. Por ejemplo, entre las mejoras de los servicios operativos figura una herramienta diseñada para predecir el riesgo de que los jóvenes de 14 años no cuenten con ninguna forma empleo, educación u otro tipo de formación a la edad de 18 años. La herramienta permite una intervención temprana y selectiva en las escuelas para reducir el riesgo de obtener este resultado.

Fuente: Mikhaylov, S.J., M. Esteve and A. Campion (2018), "Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration", *Phil. Trans. R. Soc.* A376: 20170357, http://dx.doi.org/10.1098/rsta.2017.0357.

Diseñar procesos eficaces de adquisición de IA en el sector público

En muchos casos, el talento interno y la colaboración intersectorial no serán suficientes. Será necesario que los gobiernos adquieran conocimientos y capacidades del sector privado mediante procesos de adquisición pública. Dada la incertidumbre del campo de la IA y la falta de mercados y normas preparados en la actualidad, podría ser difícil redactar contratos detallados que equilibren la obtención de servicios y la reducción de riesgos. Es poco probable que funcione una adquisición en condiciones de mercado en la que las empresas presten servicios para el gobierno de conformidad con contratos legales y requisitos técnicos detallados, y podría ser necesario que los gobiernos establezcan relaciones de colaboración a más largo plazo con los socios de la prestación de servicios. Tal vez les sea más conveniente adoptar estrategias de adquisición innovadoras para fomentar la innovación y la creación de mercados muy activos y competitivos para los bienes y servicios de IA. En la Figura 4.5 se muestra el grado en el que la adquisición pública de productos de tecnología avanzada en los países de la OCDE toma en cuenta la innovación, así como el precio.

Figura 4.5: Adquisición pública de productos de tecnología avanzada en los países de la OCDE



Nota: La cifra se basa en las respuestas a la pregunta “En su país, ¿en qué medida las decisiones sobre adquisiciones públicas fomentan la innovación?” 1 = para nada; 7 = en gran medida].
Fuente: WEF Executive Opinion Survey 2015 (encuesta a más de 14,000 ejecutivos de negocios de más de 140 países) http://reports.weforum.org/global-information-technology-report-2016/networkedreadiness-index; Oxford Insights Government AI Readiness Index, 2019, www.oxfordinsights.com/aireadiness2019.

Si no se promueve la diversidad, la apertura y normas ética y técnicamente sólidas en la adquisición de servicios de IA, podrían seguirse rumbos no óptimos para la IA que afiancen el poder de mercado de las grandes empresas, limiten la rendición de cuentas y socaven los valores sociales (Mateos-Garcia, 2018). Las leyes en materia de propiedad intelectual y demás normas que protegen el software patentado pueden hacer que “los sistemas sean

opacos e incomprensibles, lo que dificulta la evaluación de los sesgos, la impugnación de las decisiones y la corrección de los errores". Con la finalidad de mantener la confianza del público y hacer frente a las asimetrías de la información, las empresas que prestan servicios y bienes de inteligencia artificial deben estar sujetas a normas estrictas de responsabilidad, transparencia, equidad y privacidad. El Gobierno de Canadá ha elaborado una "lista de proveedores" (Recuadro 4.28) para ayudar a las oficinas gubernamentales a agilizar sus adquisiciones y seleccionar a los proveedores con experiencia en los aspectos éticos de la IA[37] (véase el caso de estudio del Anexo A). Si bien no siempre se exige a los proveedores externos que hagan público su software de propiedad exclusiva, los gobiernos deberían incorporar requisitos que les permitan acceder al código fuente para fines de auditoría a fin de comprender por qué se tomaron ciertas decisiones importantes.

Recuadro 4.28: La Lista de Proveedores de IA [*AI Source List*] del Gobierno de Canadá para el fomento de adquisiciones innovadoras

El Gobierno de Canadá creó una Lista de Proveedores de IA con 73 distribuidores preaprobados "para proporcionar a Canadá servicios, soluciones y productos de IA responsables y eficaces". Al tomar esta lista como base, las dependencias gubernamentales pueden agilizar la adquisición a empresas que han demostrado tener la capacidad de proporcionar bienes y prestar servicios de IA de calidad.

Para poder figurar en la lista, se exige a los proveedores que demuestren su competencia en los aspectos éticos de la IA, así como en la implementación y el acceso a profesionistas de primera categoría. Las empresas que respondieron a la "Invitación para ser incluido en la lista" tuvieron que demostrar a un panel interdisciplinario que cumplían con estos requisitos. La lista cuenta con tres niveles, cuyos requisitos son cada vez más exigentes a medida que aumentan. El nivel más bajo tiene requisitos menos estrictos, lo que facilita la inclusión de las pequeñas empresas de reciente creación, impulsando así la innovación y la creación de un mercado mucho más activo.

Asimismo, ésta apoya la innovación iterativa e impulsada por misiones al permitir que se encargue a múltiples empresas el desarrollo de servicios de las primeras etapas para tratar un problema. Esto permite un intercambio eficaz de información y una aproximación ágil para mitigar la incertidumbre de aquellas aproximaciones con el potencial de ser disruptivas.

El proceso de elaborar y mantener esta lista de proveedores de servicios de IA también es una forma importante en que el Gobierno de Canadá establece relaciones a más largo plazo con empresas privadas. Este diálogo facilita el desarrollo de expectativas compartidas y la comprensión mutua de los desafíos que pueden estar enfrentando, y que son relevantes para los organismos del sector público.

Fuente: https://buyandsell.gc.ca/procurement-data/tender-notice/PW-EE-017-34526; https://buyandsell.gc.ca/cds/public/2018/09/21/5e886991ecc74498b76e3c59a6777cb6/ABES.PROD.PW__EE.B017.E33817.EBSU001.PDF.

[37] www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

Considerar las necesidades de infraestructura

El tema del que nadie quiere hablar cuando se debate sobre el uso de la IA en el sector público es la necesidad de una infraestructura sólida, robusta y flexible. Si bien los datos y los algoritmos son importantes cuando se trata de la implementación de proyectos de IA, también se debe cumplir con requisitos importantes de infraestructura para poder llevar a cabo estos proyectos. Las tecnologías e infraestructuras del pasado no suelen ser adecuadas para permitir el uso de tecnologías y técnicas disruptivas, como el aprendizaje automático. Sin embargo, actualmente los gobiernos suelen tener dificultades para encontrar un equilibrio entre mantener la infraestructura de TIC heredada, la cual es antigua y obsoleta y, al mismo tiempo, tratar de ser innovadores, lo cual requiere la construcción, el desarrollo o la adquisición de nuevos equipos e infraestructuras de TIC (Desouza, 2018). En conversaciones con funcionarios gubernamentales, el OPSI ha observado que, en muchos casos, a los gobiernos se les dificulta la adopción de tecnologías de eficacia comprobada, como el cómputo en la nube, que son importantes para el progreso de la IA.

Cuando se consideran temas como la exploración y adopción de la IA en el sector público, los gobiernos deben tener en cuenta la infraestructura que actualmente está disponible y las futuras ambiciones tecnológicas del gobierno. Para hacerlo, deben incluir inversiones en infraestructura y mantenimiento como partes medulares de este plan. En lo que respecta específicamente al aprendizaje automático, se recomienda a los gobiernos prestar especial atención a la infraestructura de gestión y almacenamiento de datos, por ejemplo, explorando HADOOP o Spark, y a los requisitos de hardware para ejecutar los algoritmos, por ejemplo, obteniendo potentes grupos de CPU y GPU. Una segunda pregunta para explorar las diversas estrategias, como la infraestructura local frente a la infraestructura pública basada en la nube y las soluciones de nube híbridas. Esta decisión debe tomarse teniendo en cuenta los riesgos de seguridad, la privacidad de los datos (por ejemplo, dónde está permitido almacenarlos) y el costo.

El tema de la infraestructura y cómo superar la tecnología heredada es inmenso. Es una conversación que va mucho más allá de la IA, y es un tema demasiado profundo como para que el OPSI pudiera cubrirla por completo. Varios países están adoptando una serie de estrategias para hacer frente a los problemas relacionados con la tecnología heredada, como el examen del Fondo de Modernización Tecnológica (TMF) del Gobierno de los Estados Unidos que se menciona en el Recuadro 4.7 del presente capítulo. También se dispone de una serie de recursos que pueden ayudar a los funcionarios públicos a comprender mejor las necesidades de infraestructura, por ejemplo:

- *Implementación de IA #2: Perspectivas de la infraestructura de la IA* [*Implementing AI #2: Perspectives from AI Infrastructure*] (entrada de blog) [38]
- *¿Quieres usar la IA y el aprendizaje automático? Necesitas la infraestructura adecuada* [*Want to use AI and machine learning? You need the right infrastructure*] (artículo) [39]
- *Explorando las necesidades de infraestructura de la IA* [*Exploring the infrastructure needs of AI*] (artículo)[40]

38 https://medium.com/datadriveninvestor/implementing-ai-2-perspectives-from-ai-infrastructure-cc9a3f49569c.
39 www.networkworld.com/article/3329861/want-to-use-ai-and-machine-learning-you-need-the-right-infrastructure.html.
40 www.zdnet.com/article/exploring-the-infrastructure-needs-of-ai.

Reconocer el potencial de cambios futuros significativos y prepararse mediante la innovación anticipada

En el presente informe se ha examinado la naturaleza de la IA, los diferentes enfoques y los ingredientes necesarios, su utilización en los gobiernos y los usos que éstos le dan, y se ha estudiado y proporcionado orientación sobre algunas de las consecuencias de su implementación en el sector público.

Como se ha demostrado, a pesar de la larga historia de la IA y las opiniones que la rodean, todavía hay mucho que aprender sobre esta tecnología, y gran parte de cómo evolucionará sigue siendo algo desconocido. La IA presenta una oportunidad estelar para el gobierno, sin embargo, los gobiernos podrían tener que enfrentarse a desafíos técnicos y problemas de implementación que les impedirían aprovechar todo su potencial. También hay varias incógnitas importantes que sólo se resolverán con el tiempo a medida que se desarrolle la tecnología y se estudien y exploren sus posibles usos.

Esperar a que se resuelvan esas incógnitas es un lujo que la mayoría de los gobiernos no pueden permitirse. Esperar significa que otros elegirán la tecnología, las normas, las expectativas y las vías de desarrollo. Al mismo tiempo, también significa que otros tomarán decisiones con respecto a las competencias, capacidades, herramientas e infraestructura, inversiones y lecciones, beneficios y fracasos, de los cuales se puede aprender mucho. Esperar significará convertirse en un agente pasivo que sólo toma la tecnología de los demás en lugar de un formador de opciones, algo que podría tener costos y desventajas importantes, no sólo desde el punto de vista de la eficacia, sino también en términos de seguridad, desarrollo industrial, valores nacionales y otras preocupaciones gubernamentales fundamentales. El dilema de Collingridge ayuda a ilustrar este punto (Recuadro 4.29).

Recuadro 4.29: El dilema de Collingridge

"Cuando aún es fácil cambiar una tecnología, al principio de su uso, se sabe tan poco sobre las posibles consecuencias que traerá que los actores sociales no pueden anticiparse adecuadamente a ellas. Por lo tanto, no les es posible evitar ninguna consecuencia no deseada en esta etapa mediante la adaptación de la tecnología. Una vez que una tecnología se utiliza extensamente, las consecuencias se conocen, pero en esa etapa es muy difícil adaptar la tecnología para evitar efectos indeseados porque para ese momento ésta ya se ha arraigado en la sociedad".

Existen dos soluciones: "una es aumentar el conocimiento en las etapas iniciales del desarrollo de una nueva tecnología y la otra es aumentar el control social sobre las trayectorias tecnológicas".

Fuente: Mikhaylov, S.J., M. Esteve, A. Campion (2018), "Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration", *Phil. Trans. R. Soc.* A376: 20170357, http://dx.doi.org/10.1098/rsta.2017.0357.

Así pues, la IA será una aventura llena de sorpresas, tanto bienvenidas como indeseadas, y de acontecimientos inesperados e imprevistos. En la siguiente sección se exploran algunos aspectos clave de este futuro complejo, y algunos métodos que pueden ayudar a guiar a aquellas personas que se dedican a esta tecnología.

La IA introducirá el conocimiento no humano

En el pasado, la humanidad se ha visto limitada por el conocimiento que ha sido capaz de crear o asimilar. Esto incluye copiar o inspirarse en la naturaleza. Si bien la base de conocimientos de la que disponen los seres humanos ha tomado giros en ocasiones sorprendentes, no ha dejado de ser conocimiento humano, afianzado en la experiencia y el contexto humanos y derivado de estos.

La IA promete cambiar esto, quizás a niveles fundamentales. Ya existen casos de IA que crea nuevos conocimientos que actualmente son inexplicables para el pensamiento humano (Recuadro 4.30).

Recuadro 4.30: La IA reinventa uno de los juegos más antiguos de la humanidad

El juego Go es complejo e implica un grado elevado de estrategia. Aunque los movimientos básicos son simples, si se toma en cuenta que cada jugador coloca piedras de diferentes colores en una cuadrícula, se ha estimado que el límite mínimo de posibles posiciones en el tablero es de 2 x 10^{170} (Wikipedia). Para poner esta cifra en contexto, sólo se estima que hay 10^{80} átomos en todo el universo. Por lo tanto, no es un juego en el que se gane fácilmente mediante la “fuerza bruta” del cómputo o el poder de pensamiento, al igual que en el ajedrez. Dominar el juego requiere una profunda comprensión de las tácticas y estrategias.

DeepMind, una de las empresas de Alphabet, ha desarrollado un programa informático que combina un árbol de búsqueda avanzada con redes neuronales profundas. Su entrenamiento empezó con juegos de Go previamente disputados para ayudarle a aprender las reglas y la forma de juego, y luego diferentes versiones del programa jugaron entre sí, lo que le ayudó a aprender de la experiencia a través del aprendizaje por refuerzo.

En 2016, AlphaGo jugó contra el campeón humano de ese entonces y ganó 4-1. Poco después se publicó una versión mejorada titulada AlphaGo Zero. La característica distintiva de esta versión era que el sistema aprendía simplemente jugando contra sí mismo, empezando con un juego completamente aleatorio, en lugar de utilizar juegos jugados por humanos como referencia, con el único límite de las reglas básicas del juego.

AlphaGo Zero superó incluso a AlphaGo, venciéndolo por 100 juegos a cero. Algunos de los movimientos realizados por AlphaGo, y su sucesor Zero, han sido descritos como incomprensibles, inhumanos y “extraterrestres”. En un juego que ha existido durante 2,500 años, una máquina vio posibilidades y tácticas nunca percibidas por ningún hábil jugador humano. Estas ideas han cambiado para siempre la forma en que los humanos jugarán el juego.

Fuente: https://deepmind.com/research/case-studies/alphago-the-story-so-far;
https://en.wikipedia.org/wiki/Go_(game);
www.theatlantic.com/technology/archive/2017/10/alphago-zero-the-ai-that-taught-itself-go/543450.

Si bien los humanos inventaron la IA, y por lo tanto la IA refleja la humanidad (junto con sus fortalezas y prejuicios), es la primera de las herramientas de la humanidad que tiene el poder y la capacidad de desviarse de las ideas preconcebidas y supuestos de la humanidad, y de crear ideas y conocimientos que son fundamental y verdaderamente extraños, en otras palabras, no sólo novedosos sino también separados y diferentes de las formas de pensar del ser humano. La IA no está necesariamente ligada a los conocimientos, expectativas, tradiciones, normas, valores o creencias humanas. Por primera vez, el conocimiento puede originarse desde fuera de las percepciones generadas u originadas biológicamente. El significado de este hito sólo se comprenderá plenamente en retrospectiva (si es que alguna vez sucede). Lo que se puede decir con cierto grado de confianza es que esto dará paso a posibilidades nunca antes vistas.

La IA hace que el futuro sea más desconocido, pero también más modelable

El mundo actual se ha descrito en términos de un estado de volatilidad, incertidumbre, complejidad y ambigüedad (VUCA, por sus siglas en inglés). Es imposible saber lo que deparará el futuro, a pesar del deseo de previsibilidad, estabilidad y certeza.

Esto se verá exacerbado por la IA. Como se describe en el presente informe, la IA introducirá nuevas y más potentes capacidades, probablemente a un costo relativamente bajo y a disposición no sólo de los gobiernos y las grandes empresas, sino también de los consumidores individuales. A medida que más personas están facultadas con mayores capacidades, las posibilidades que depara el futuro se diversifican cada vez más. Más gente capaz de hacer más cosas a un costo menor equivale a un mayor rango de escenarios futuros posibles.

Esta situación se magnificará aún más cuando la IA empiece a introducir nuevos conocimientos y tácticas, imprevistas hasta ahora para el ser humano. Un futuro determinado únicamente por motivaciones, intereses y conocimientos puramente humanos ya era bastante difícil de afrontar, y la adición de la IA hace que esto sea inconcebible.

Al mismo tiempo, la IA actualmente permite, y posibilitará aún más, la elaboración de modelos y simulaciones del futuro más potentes. Por ejemplo, el aprendizaje automático puede ayudar a enfrentar el cambio climático al posibilitar que se hagan mejores predicciones climáticas, ilustrar los efectos del cambio climático y ayudar a capacitar a los agentes para preparar mediciones más detalladas (Snow, 2019). Ya sea en forma de modelos climáticos y meteorológicos, o a través del modelado de marcos hipotéticos y tendencias sociales, económicas o medioambientales, la IA podría ofrecer herramientas mucho más poderosas para ayudar a dar sentido a patrones que los humanos no han podido percibir, reconocer o conceptualizar anteriormente debido a su tamaño.

Por lo tanto, es probable que la IA contribuya a que se creen nuevas e incalculables posibilidades, además de ofrecer herramientas y medios para ayudar a comprenderlas más rápidamente. Los humanos tendrán más información que nunca antes y al mismo tiempo, tendrán menos certeza sobre lo que sucederá.

Un futuro en constante cambio requiere aprender a no descartar ninguna opción

Si hay múltiples futuros posibles, pero no se sabe con certeza cuál de ellos ocurrirá o incluso si uno pudiera ser más preferible o deseable, entonces es arriesgado invertir de

forma excesiva en cualquier conjunto de supuestos sobre el futuro. En un contexto de gran incertidumbre, resulta conveniente mantener activa una variedad de opciones diferentes, de modo que sea posible cambiar o pasar a otras alternativas a medida que se aprende más sobre lo que se necesita.

Esto sugiere que los gobiernos deben adoptar un enfoque ágil y adaptable a fin de ajustarse a las nuevas oportunidades y a los cambios de comportamiento. Dado que no existe un límite para el desarrollo de la inteligencia de la IA, muchas de las tareas que ésta no puede llevar a cabo de manera eficaz hoy en día serán factibles en el futuro. Las estrategias y marcos en torno a la IA deben ser lo suficientemente flexibles para evolucionar con las capacidades y contextos cambiantes. La tecnología basada en IA es dinámica y la forma en que interactúa con los humanos en los complejos sistemas de prestación de servicios públicos evolucionará con el tiempo (Kattel, 2019).

Por ejemplo, en términos prácticos esto sugiere que los gobiernos deben evitar los contratos a largo plazo que limiten al sector público a soluciones y formas de trabajo actuales o exclusivas. El Manual de Servicio [*Service Manual*] del Gobierno del Reino Unido sostiene que al elegir una tecnología, "lo más importante es tomar decisiones que le permitan: cambiar de opinión en una etapa posterior [y] adaptar su tecnología a medida que cambie su comprensión de cómo satisfacer las necesidades del usuario".[41] Del mismo modo, los planes estratégicos deben ser documentos vivos y no artefactos estáticos, cuya actualización implique exámenes periódicos para monitorear la aplicación y evaluar si las hipótesis de planificación siguen siendo válidas. Este punto de vista es la idea principal de *Inteligencia Artificial: Una visión estratégica para Luxemburgo* [*Artificial Intelligence: A Strategic Vision for Luxembourg*],[42] una estrategia viva puesta en marcha por el Gobierno de Luxemburgo que se actualizará regularmente. En Francia, el Etalab del Gobierno ha elaborado directrices sobre el uso de algoritmos para las administraciones públicas (véase el Recuadro 4.15), con una versión editable de éstas que se almacena en GitHub, donde los colaboradores pueden modificarlas para mejorar el contenido.[43] El OPSI ha observado una tendencia a la alza de este tipo de directrices vivas en el sector público, las cuales se adaptan bien al carácter dinámico y no estático de la evolución de la IA y de las tecnologías basadas en la IA (OCDE, 2017b).

Esto también sugiere que los gobiernos deben mejorar su capacidad para captar incluso señales débiles, que indiquen cómo puede evolucionar el futuro desde una etapa más temprana. Esto les permitirá conocer los lugares y momentos óptimos para intervenir, sin esperar a que los procesos y tendencias se arraiguen y, por lo tanto, cambiarlos se vuelva más costoso y difícil. La IA puede impulsar mejores formas de dar sentido a los futuros y poner a prueba situaciones hipotéticas y supuestos.

41 www.gov.uk/service-manual/technology/choosing-technology-an-introduction.
42 https://digital-luxembourg.public.lu/initiatives/artificial-intelligence-strategic-vision-luxembourg.
43 www.etalab.gouv.fr/algorithmes-publics-etalab-publie-un-guide-a-lusage-des-administrations.

Recuadro 4.31: IBM y la Previsión Automática [*Machine Foresight*]

IBM define la Previsión Automática "como la tecnología y el fundamento teórico para la generación de tales situaciones hipotéticas factibles. El conocimiento obtenido de estas situaciones equipa a los encargados de la toma de decisiones, con información que pueden utilizar para inclinar el futuro hacia estados deseables y alejarlo de situaciones no deseadas mediante la adopción de las medidas adecuadas". Esto puede incluir:

- **planificación de situaciones hipotéticas**, mediante la cual las noticias y los datos de las redes sociales se utilizan para identificar historias importantes para el usuario final, y luego se aplican algoritmos de planificación de IA para generar situaciones hipotéticas futuras
- **plataformas de simulación,** p. ej., utilizar la plataforma de simulación empresarial para ayudar a los funcionarios gubernamentales de un país determinado a simular determinadas situaciones hipotéticas y un conjunto de medidas como respuesta a éstas.
- **descubrimiento avanzado**, en el que el uso de la IA puede ayudar a identificar oportunidades inesperadas de innovación y determinar dónde puede tener más éxito la inversión.

Fuente: www.ibm.com/blogs/nordic-msp/ai-machine-foresight.

El futuro del trabajo como ejemplo

A medida que los robots, la Inteligencia Artificial y la transformación digital van permeando cada vez más el mundo laboral, la OCDE estima que el 14% de los empleos de los países miembros corren un alto riesgo de volverse automatizados, lo cual cambiará radicalmente las tareas que deben realizarse en el 32% de los empleos (OCDE, 2019b). Nadie sabe hasta qué punto estas estimaciones son precisas, pero la mayoría coincide en que habrá ramificaciones masivas de la IA y la automatización en el futuro laboral y, por ende, en los medios de subsistencia y la razón de ser de los seres humanos.

Mientras que la OCDE y otros han ofrecido reflexiones y estimaciones, existe una amplia incertidumbre sobre cómo se desarrollarán exactamente las cosas. Debido a que los posibles costos y beneficios sociales y económicos que están en juego son enormes, los gobiernos no pueden simplemente adoptar una actitud pasiva y esperar a ver lo que sucede. Necesitan comprometerse con las posibilidades de manera continua, utilizando una variedad de métodos para probar las suposiciones, extrapolar los posibles futuros y sus implicaciones, y pensar en qué futuros se desean realmente.

La IA y la humanidad: Una relación llena de potencial (riesgos y beneficios)

> *"La búsqueda de diferentes tipos de futuros se topa directamente con el obstáculo de que, por definición, el futuro no puede existir en el presente, porque si así fuera ya no sería el futuro, sino el presente. Y, sin embargo, como todos saben, el futuro juega un papel en el presente. ¿Cómo puede afectarnos de esta forma algo que no existe?" (Riel Miller, 2018)*

Aunque no hay certidumbre sobre el futuro, eso no significa que las expectativas del futuro no tendrán efecto sobre las medidas que se tomen ahora. La IA tendrá un gran impacto en el mundo; la pregunta es "¿de qué tipo y qué se debe hacer al respecto?" Las suposiciones sobre el futuro y las medidas adoptadas darán forma al futuro que viene.

En este manual introductorio se han expuesto algunos de los antecedentes y fundamentos técnicos de la IA, así como el panorama actual y los desafíos e implicaciones que los funcionarios públicos deben afrontar al día de hoy. Al pensar en el futuro, independientemente del futuro de la IA, el sector público debe comprometerse a ayudar a darle forma y a marcar el camino a seguir. La IA tendrá efectos en todos los ámbitos de la sociedad, en la economía e incluso en el medio ambiente. El sector público no puede permanecer como un espectador. En este manual introductorio se ha intentado ayudar a los funcionarios públicos a comprender mejor los problemas, las oportunidades y las posibles consecuencias, así como dar algunas orientaciones prácticas sobre lo que se puede hacer *ahora*. Sin embargo, también pueden tomar medidas específicas para ayudar a mejorar las cosas *en el futuro*.

En el futuro, los gobiernos deberán evaluar de forma continua los temas centrales de la IA desde una perspectiva sistémica, teniendo en cuenta la forma en que las diferentes organizaciones, sectores e interesados se relacionan con la IA y se ven afectados por ella en la actualidad, así como la forma en que pueden relacionarse y verse afectados por la IA en el futuro. Esto requiere una planificación a largo plazo en torno a futuros e hipótesis estratégicas que podrían ayudar a tener en cuenta los posibles riesgos con antelación, pero también cambiar dinámicamente la postura estratégica con base en el desarrollo tecnológico. El OPSI considera que la mejor manera en que el gobierno puede participar en este proceso y asumir la evaluación continua es participar en el proceso de innovación, es decir, crear capacidades de anticipación y gestión de la innovación anticipada en el ámbito de la IA (véanse las definiciones y los conceptos en el Recuadro 4.32).

Recuadro 4.32: Definiciones y conceptos de la Gestión de la Innovación Anticipada

- *Anticipación* es el proceso de creación de conocimiento, no importa cuán tentativo o cualificado sea, sobre diferentes futuros posibles. Esto puede incluir, entre otras cosas, el desarrollo no sólo de hipótesis sobre alternativas tecnológicas, sino también de situaciones hipotéticas tecno-morales (basadas en valores) del futuro (Normann, 2014).
- *Gestión anticipada* es el proceso de actuar sobre una variedad de aportaciones para gestionar las tecnologías emergentes basadas en el conocimiento y los acontecimientos socioeconómicos, mientras que dicha gestión todavía es posible (Guston, 2014). Esto puede implicar aportaciones de diversas funciones de gobierno (previsión, participación, elaboración de políticas, financiamiento, regulación, etc.) de manera coordinada.
- *Regulación anticipada* es una función de la gestión anticipada que utiliza medios normativos para crear espacios aislados, demostradores, bancos de pruebas, etc. para que surjan diversas opciones tecnológicas. Esto requiere un desarrollo iterativo de los reglamentos y las normas en torno a un campo emergente (Armstrong y Rae, 2017).

Recuadro 4.32: Definiciones y conceptos de la Gestión de la Innovación Anticipada *(Cont.)*

- *Gestión de la innovación anticipada* es una capacidad generalizada para explorar de forma activa las opciones como parte de una gestión anticipada más amplia, con el objetivo particular de estimular las innovaciones (productos, servicios y procesos novedosos en el contexto, aplicados y que cambian de valor) relacionadas con futuros inciertos con la esperanza de configurar los antedichos mediante prácticas innovadoras (OPSI, 2019).

Fuente: OPSI (2019), Anticipatory Innovation Governance Expert Group Meeting, Presentation, 30 May; Guston, D.H. (2014), "Understanding 'anticipatory governance'", *Social Studies of Science*, Vol. 44/2, pp. 218-242; Nordmann, A. (2014), "Responsible innovation, the art and craft of anticipation", *Journal of Responsible Innovation*, Vol. 1/1, pp. 87-98; Armstrong, H. and J. Rae (2017), *A Working Model for Anticipatory Regulation*, Nesta, London.

Anticiparse no significa predecir el futuro, sino formular preguntas sobre futuros factibles y luego actuar en consecuencia mediante la creación de espacio para la innovación (p. ej., mediante la reglamentación) o creando los mecanismos para explorar diferentes opciones dentro del propio gobierno. La mayoría de los gobiernos no disponen actualmente de un sistema de gestión de la innovación anticipada (esos mecanismos suelen estar aislados en determinados ámbitos o funciones de política, como la previsión). Con ese fin, el OPSI, junto con varios países de la OCDE, ha puesto en marcha recientemente un proyecto de Gestión de la Innovación Anticipada para poner a prueba en la práctica, diferentes mecanismos de anticipación y ofrecer a los gobiernos, orientación y recomendaciones sobre este tema. Los resultados de este trabajo se publicarán en el sitio web del OPSI[44] y se compartirán con los suscriptores del boletín del OPSI[45]

Reunir todos los ingredientes: Un marco para que los gobiernos desarrollen su estrategia de IA

No se puede esperar que este informe capture o aborde de forma suficiente todas las oportunidades e implicaciones que vienen con la IA. La IA es un tema enorme. La IA, hasta cierta medida, alterará nuestras percepciones de lo que significa ser humano. Al igual que en el juego Go, la IA nos enseñará cómo pensamos. Al igual que en el ejemplo de los algoritmos que se utilizaron en el sistema de justicia penal de los Estados Unidos (Recuadro 4.17), la IA nos enseñará sobre nuestros sesgos y nuestras relaciones. Puede cambiar la relación que tenemos con nuestro mundo y la forma en que interactuamos con él, al darnos nuevas perspectivas, nuevas herramientas para medir y entender el mundo, e incluso nuevos conocimientos y perspicacia. La relación entre el ser humano y la IA evolucionará con el tiempo. El sector público tendrá un papel crucial para asegurar que esa relación realce las partes de la humanidad que son más preciadas.

44 https://oecd-opsi.org.
45 https://oecd-opsi.org/newsletter.

En esta sección se reúnen las implicaciones y consideraciones expuestas anteriormente, y las sitúa en un marco de alto nivel para ayudar a los gobiernos a pensar en su estrategia de IA para transformar los servicios públicos y la administración.[46]

Se recomienda a los gobiernos que adopten diferentes estrategias basadas en su contexto estratégico, sus prioridades y sus capacidades iniciales. Una estrategia de IA debe incluir lo siguiente:

- **Capacidades iniciales**: una evaluación de la situación estratégica actual de la dependencia y los desafíos que la IA podría ayudar a resolver.
- **Objetivos**: qué es lo que la dependencia desea lograr con la IA y los principios que sustentarán las medidas que se tomen para conseguirlos.
- **Estrategias**: las acciones concretas que se llevarán a cabo para lograr estos objetivos.

En la sección que figura a continuación se exponen los elementos que los gobiernos deberían considerar incluir en su estrategia de IA. Para que una estrategia sea eficaz, será necesario obtener un panorama claro de la situación actual a través de la supervisión. No debe considerarse que los elementos de la estrategia de IA son secuenciales o lineales. Sin embargo, tendrán que desarrollarse simultáneamente y la estrategia deberá ser un documento vivo que pueda ser iterado para garantizar su coherencia y adaptarse a medida que evolucione el contexto.

[46] El presente capítulo se basa en una serie de fuentes, en particular www.nesta.org.uk/blog/10-questions-ai-public-sector-algorithmic-decision-making/; https://hbr.org/2018/04/a-simple-tool-to-start-making-decisions-with-the-help-of-ai; https://docs.google.com/presentation/d/1TAJ2A4NvMLFi7b0mTvNyL1pMVRy84UhzhgcsXknhR2g/edit#slide=id.p1); Agarwal, A., J. Gans, J. and A. Goldfarb (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business Press, https://faculty.ai/products-services/ai-strategy.

Marco para elaborar una estrategia de IA

Capacidades iniciales	Objetivos	Estrategias
Determinar las fortalezas y debilidades actuales mediante el mapeo: ● Capacidades internas para IA. ● Activos de datos gubernamentales y externos. ● Proyectos gubernamentales existentes sobre IA y ciencia de datos. *Evaluar el contexto estratégico:* ● Actitudes del público y de la fuerza de trabajo hacia el gobierno y la IA. ● Marco jurídico y normativo actual. ● Compromisos gubernamentales e internacionales existentes, instituciones. ● Los conocimientos especializados del sector académico y privado que podrían aprovecharse. *Identificar problemas operativos específicos que la IA podría ayudar a resolver:* ● Adoptar un enfoque multidisciplinario y diverso para decidir si la IA es la mejor solución a un problema de políticas y, en caso afirmativo, cómo debe diseñarse y aplicarse una estrategia basada en el mismo. ● Poner en marcha procesos para comprender las necesidades de los usuarios. ● Crear mecanismos para adecuar los recursos a los problemas prioritarios. ● Definir la decisión específica que la IA tomará o apoyará. ● Considerar quiénes se verán afectados por esta decisión y los riesgos asociados en caso de fracaso. ● Explorar las formas en que el servicio tendrá que ser rediseñado para aprovechar el impacto de la IA.	*Decidir cuáles serán los objetivos que la IA ayudará al gobierno a lograr:* ● Determinar las formas en que la IA generará valor público y especificar las misiones en las que la IA puede ser parte de la solución. ● Involucrar a las partes interesadas al definir los objetivos. ● Hacer espacio para la experimentación y el aprendizaje. *Definir y comunicar a las partes interesadas los principios que conformarán la forma en que la IA se utiliza en el gobierno:* ● Equidad e imparcialidad. ● Transparencia y responsabilidad. ● Privacidad y autonomía individual.	*Garantizar el acceso del gobierno a personal capacitado para el uso de la IA:* ● Crear reservas de talentos y elaborar planes de contratación y retención de expertos técnicos internos. ● Aprovechar los conocimientos especializados externos mediante asociaciones y colaboración. ● Diseñar procesos eficaces de adquisición de IA en el sector público. ● Crear un cuadro de funcionarios públicos que entiendan las cuestiones legales, éticas, técnicas y de gestión en torno a la IA. ● Establecer planes de financiamiento y garantizar la disponibilidad de recursos en los planes fiscales del gobierno. *Garantizar el acceso ético a datos e infraestructuras de calidad y su utilización:* ● Determinar qué datos se necesitan para atender los problemas. ● Decidir las maneras para obtener datos de entrada de suficiente calidad y que sean suficientemente representativos de la población objetivo para hacer predicciones precisas con un sesgo mínimo. ● Desarrollar una estrategia de datos que cumpla con la ley de protección de datos y las mejores prácticas, y que sea compatible con los principios de la IA. ● Asegurar que se disponga de una infraestructura de datos importante (p. ej., servicios de nube híbridos). *Establecer marcos jurídicos, éticos y técnicos para poner los principios en práctica:* ● Vigilar el cumplimiento de los principios durante la implementación para dar un seguimiento a los progresos, e identificar y responder a las cuestiones que surjan. ● Poner en marcha salvaguardas contra los sesgos y la injusticia. ● Aclarar el papel adecuado de los humanos en el proceso de toma de decisiones. ● Desarrollar estructuras de responsabilidad abiertas y transparentes. *Prepararse para cambios futuros:* ● Garantizar la apertura y la flexibilidad en los planes y contratos futuros. ● Seguir los conceptos de Gestión de Innovación Anticipada del OPSI.

Anexo A: Casos de estudio

Como puede verse a lo largo de este manual, los gobiernos están asumiendo un papel cada vez más activo en lo que respecta al diseño y la ejecución de proyectos de IA, así como en la creación de las condiciones propicias y la orientación necesarias para garantizar que los proyectos se implementen de manera eficiente, eficaz y ética. En esta sección se presentan varios casos de estudio que ilustran los métodos que los gobiernos están utilizando para lograrlo, a fin de crear nuevas políticas y servicios innovadores. Se incluyen casos sobre proyectos específicos de IA, así como métodos y marcos más amplios para considerar la aplicación de IA. Esta recopilación de casos no es exhaustiva, pero ayuda a crear un conjunto de conocimientos sobre las diferentes prácticas implementadas en todo el mundo, el contexto en el que surgieron, las tecnologías y los enfoques utilizados y, cuando es posible, las lecciones aprendidas de estas experiencias.

El uso de la IA para obtener información sobre la toma de decisiones públicas en Bélgica

Problema

Los gobiernos se esfuerzan cada vez más por elaborar políticas y servicios orientados a los ciudadanos. Por defecto, esto requiere una amplia colaboración con los ciudadanos y residentes a fin de comprender sus perspectivas, opiniones y necesidades. Las plataformas de participación digital son instrumentos importantes para lograr esto y mejorar la capacidad de respuesta de los gobiernos. Sin embargo, el análisis de los altos volúmenes de entradas datos de ciudadanos recopilados en estas plataformas consume mucho tiempo y es desalentador para los funcionarios gubernamentales, y les impide descubrir aportaciones valiosas. Por lo tanto, no basta con crear una plataforma de participación digital: el proceso de análisis de datos tiene que ser más accesible para que los funcionarios públicos puedan aprovechar la inteligencia colectiva y de esta forma tomen decisiones mejor informadas.

Respuesta

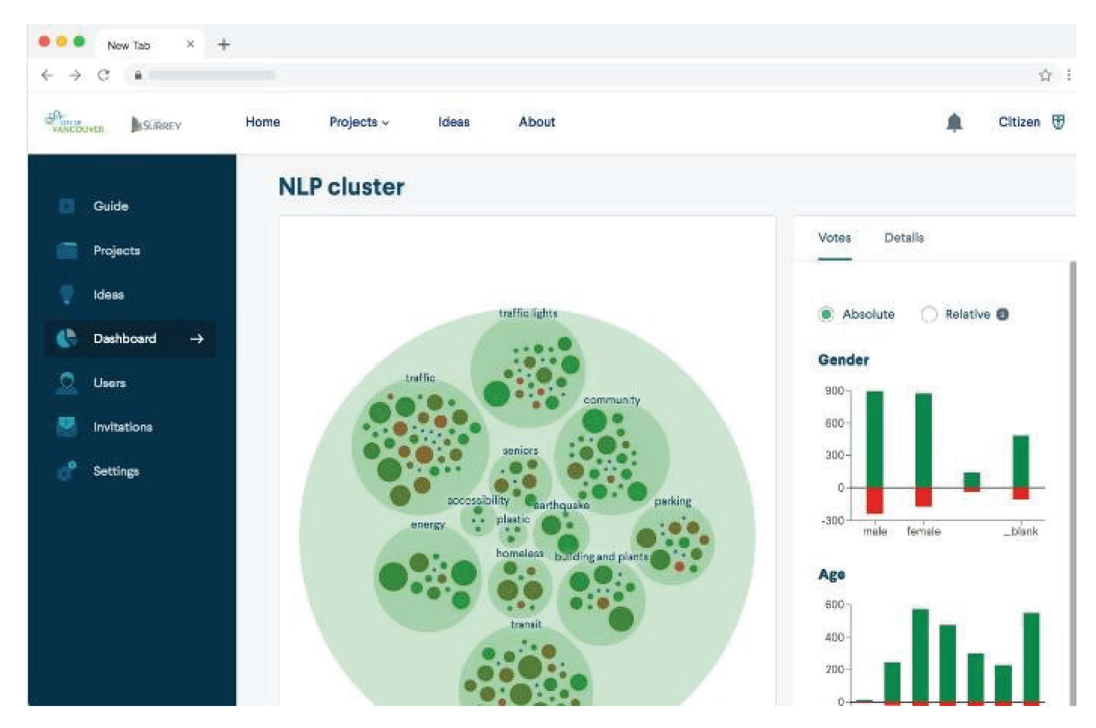
El CitizenLab de Bélgica[1] es una empresa de tecnología cívica que tiene por objeto facultar a los funcionarios públicos y proporcionarles procesos aumentados de aprendizaje automático que les ayuden a analizar las entradas de datos de los ciudadanos, tomar mejores decisiones y colaborar más eficazmente a nivel interno.[2]

[1] www.citizenlab.co.
[2] Para obtener más información sobre este caso, véase https://oecd-opsi.org/innovations/unlocking-the-potential-of-crowdsourcing-for-public-decision-makingwith-artificial-intelligence.

En cumplimiento con su misión, el CitizenLab ha desarrollado una plataforma de participación pública que utiliza algoritmos de aprendizaje automático para ayudar a los funcionarios públicos a procesar fácilmente miles de contribuciones de ciudadanos, y utilizar esos conocimientos de manera eficiente en la toma de decisiones. Los tableros de la plataforma pueden clasificar las ideas, destacar los temas emergentes, resumir las tendencias y agrupar las contribuciones similares por tema, rasgo demográfico o ubicación.

La plataforma del CitizenLab utiliza técnicas de Procesamiento del Lenguaje Natural (PLN) y de aprendizaje automático para clasificar y analizar automáticamente miles de contribuciones recopiladas en las plataformas de participación ciudadana. Los algoritmos identifican los temas principales y clasifican ideas similares en grupos, que luego pueden ser desglosados por características demográficas o ubicación geográfica.

Figura A.1: Grupo de interés comunitario



Fuente: https://oecd-opsi.org/innovations/unlocking-the-potential-of-crowdsourcing-for-public-decision-making-with-artificial-intelligence.

Los funcionarios que gestionan estas plataformas de participación ciudadana pueden acceder a esta información de un vistazo a través de tableros inteligentes en tiempo real. La función de "Modelado de temas" les permite identificar fácilmente las prioridades de los ciudadanos y tomar decisiones de forma acorde.

La plataforma permite a los funcionarios públicos desglosar los resultados por grupos demográficos y ubicación, lo que les da un mejor panorama de la variación de las prioridades. Por ejemplo, un determinado vecindario puede dar prioridad a mejores carreteras, mientras que su vecino solicita señales de alto adicionales.

En un ejemplo pertinente de principios de 2019, un número cada vez mayor de jóvenes belgas protestaba por la inacción contra el cambio climático, movimiento que se convirtió en Youth for Climate Belgium. En respuesta a ello, CitizenLab creó una plataforma de participación sobre el tema denominada Youth4Climate, e invitó a los usuarios a que presentaran ideas para hacer frente al cambio climático.[3] Durante tres meses, los usuarios enviaron 1,700 ideas, 2,600 comentarios y 32,000 votos sobre las iniciativas que querían apoyar. El sistema de inteligencia artificial analizó estos elementos, y sacó a relucir y agrupó las prioridades más importantes y que habían tenido más respaldo. El CitizenLab está utilizando las conclusiones impulsadas por la IA para elaborar un informe para los funcionarios electos con 16 recomendaciones de políticas.

A través de la iteración continua de la plataforma, CitizenLab trabaja para asegurar que los gobiernos hagan un uso óptimo de sus tableros automatizados. Además, este organismo está explorando formas en las que podría aplicar esta tecnología a conversaciones a mayor escala en redes sociales, foros públicos u otros lugares de debate en línea.

Resultados e impacto

Los gobiernos que utilizan esta plataforma han obtenido resultados positivos. La ciudad de Kortrijk, por ejemplo, utiliza los tableros inteligentes para procesar fácilmente las contribuciones de los 1,300 usuarios de su plataforma. Han agrupado las ideas de las conversaciones en temas principales y han compartido los resultados del análisis con los ciudadanos. El resultado es un diálogo real en lugar de una iniciativa de arriba hacia abajo. En otro caso, la ciudad de Temse consultó a sus ciudadanos sobre el tema de la movilidad y ubicó las ideas de la multitud en un mapa de la ciudad. Esto ayudó a la administración a identificar las zonas afectadas por cuestiones clave y a tomar decisiones sobre dónde asignar los fondos.

Al automatizar la engorrosa labor de análisis de datos, la plataforma libera tiempo para que las administraciones interactúen y participen de manera significativa con los ciudadanos. También permite a los gobiernos comprender mejor las necesidades y prioridades de los ciudadanos, lo que a su vez conduce a decisiones mejor informadas. Los gobiernos que utilizan la plataforma han reportado esos resultados.

Desde la perspectiva de los ciudadanos, este proceso abierto y transparente contribuye a fomentar la confianza y a aumentar su apoyo a las decisiones en torno a las políticas. También ha tenido un impacto positivo en la disposición de los ciudadanos a participar.

Desafíos y lecciones aprendidas

CitizenLab ha enfrentado dos desafíos principales: los algoritmos de clasificación y la adopción humana.

La plataforma utiliza un algoritmo de clasificación que agrupa, categoriza y resume las aportaciones de los ciudadanos. Debe ser fácilmente escalable, pero también debe adaptarse a los flujos de trabajo de las diferentes administraciones, ya que las taxonomías utilizadas pueden variar según el país o incluso según la región. Los algoritmos de clasificación

[3] Una descripción más detallada del proceso está disponible en www.citizenlab.co/blog/civic-engagement/youth-for-climate-case-study.

también deben aceptar múltiples idiomas en la misma plataforma y establecer vínculos semánticos entre los idiomas, lo que añade una capa adicional de complejidad técnica. Durante su trabajo en la plataforma Youth4Climate en Bruselas, CitizenLab tuvo que analizar miles de contribuciones en francés, neerlandés e inglés. Descubrieron que el mejor resultado se obtenía al traducir automáticamente los comentarios a un solo idioma y comenzar a trabajar con ellos desde ese punto.

En el lado humano, CitizenLab necesita garantizar que la tecnología responda a las necesidades reales de los usuarios para maximizar su adopción por parte de los gobiernos. El equipo ha aprendido que no se debe promocionar el producto sin antes guiar a los usuarios a través de sus beneficios. También han aprendido que la interacción hombre-máquina es crucial. El usuario debe aprender a interpretar y "confiar" en los resultados que genera la máquina y comprender el papel que estos pueden desempeñar en el flujo de trabajo diario. Los gobiernos deben considerar estas cosas antes de desplegar una solución de este tipo.

El equipo de CitizenLab también señaló varias condiciones para una implementación exitosa. La primera es promover la adopción de la plataforma. Esto implica asegurar que los funcionarios públicos entiendan sus beneficios y sientan que pueden confiar en los resultados. Explicar la metodología e integrar el proceso de participación pública con los flujos de trabajo existentes puede servir de ayuda en este aspecto. También es importante especificar una necesidad identificada, ya que el tiempo y los recursos son escasos en las administraciones, y los funcionarios públicos sólo invertirán en una herramienta si ésta tiene un valor comprobado.

En segundo lugar, el equipo constató que la calidad de las aportaciones (es decir, la retroalimentación de los ciudadanos) es fundamental para comprender con éxito las perspectivas y necesidades de los ciudadanos. Con este fin, los funcionarios públicos tienen que orientar a los ciudadanos para asegurarse de que presenten contribuciones útiles. El equipo también realizó pruebas de usuario con regularidad y perfeccionó el método con base en la retroalimentación. A nivel más estratégico, el equipo del CitizenLab descubrió que, al destacar la importancia de la participación ciudadana en los niveles más altos de la administración, se alienta a las oficinas gubernamentales y a los funcionarios públicos a buscar la opinión del público. Esto, a su vez, promovió la mejora continua de la plataforma y sus resultados.

La Estrategia Nacional de Inteligencia Artificial de Finlandia

Problema

A pesar de que es un país pequeño con 5.5 millones de habitantes, Finlandia ha declarado su intención de convertirse en líder mundial en el uso de IA. El país está bien posicionado para lograr este objetivo debido a una serie de factores. Sus ciudadanos tienen un alto nivel educativo y conocimientos tecnológicos, la economía ya presenta un amplio uso de la tecnología, el gobierno ha acumulado datos de alta calidad y, tras años de reforma, su sector público está altamente digitalizado y ha adoptado la experimentación y la innovación con los brazos abiertos. Además, las investigaciones de la empresa consultora McKinsey indican que, si Finlandia acelera el desarrollo de la IA y la automatización, puede esperar un aumento del PIB del 3% anual y un aumento neto del empleo del 5% (McKinsey & Company,

2017). Se cuenta con la combinación adecuada de facilitadores e incentivos. La pregunta clave es: ¿Qué es exactamente lo que Finlandia necesita hacer para alcanzar su potencial?

Respuesta

La era de la IA en Finlandia

En mayo de 2017, el Ministerio de Asuntos Económicos y Empleo de Finlandia creó un Programa de Inteligencia Artificial y un Grupo Coordinador para garantizar su orientación. El grupo aprovechó una amplia red de expertos para estudiar cuestiones clave sobre la mejor manera de apoyar a los sectores público y privado en la producción de innovaciones basadas en IA, la forma de posicionar los datos gubernamentales como recursos para el desarrollo económico, la forma en que la IA afectará a la sociedad y lo que el sector público debe hacer para que Finlandia avance hacia un futuro impulsado por la IA. Como consecuencia de esta labor, el Grupo Coordinador publicó dos informes clave que exponen el enfoque de Finlandia sobre la IA. *La Era de la Inteligencia Artificial de Finlandia* [*Finland's Age of Artificial Intelligence*][4] (diciembre de 2017) y *Llevando la vanguardia hacia la Era de la Inteligencia Artificial* [*Leading the Way into the Age of Artificial Intelligence*] (junio de 2019) establecen en conjunto 11 medidas clave que abarcan todos los sectores para ayudar a Finlandia a alcanzar su ambicioso objetivo:

1. Aumentar la competitividad de las empresas mediante el uso de la inteligencia artificial.
2. Utilizar eficazmente los datos en todos los sectores.
3. Garantizar que la adopción de la IA sea más fácil y rápida.
4. Asegurar el máximo nivel de conocimientos y atraer a los principales expertos en la materia.
5. Tomar decisiones y hacer inversiones audaces.
6. Crear los mejores servicios públicos del mundo.
7. Establecer nuevos modelos de colaboración.
8. Hacer que Finlandia sea un líder en la era de la IA.
9. Prepararse para que la Inteligencia Artificial cambie la naturaleza del trabajo.
10. Coordinar el desarrollo de la IA en una dirección basada en la confianza y centrada en el ser humano.
11. Prepararse para los desafíos de seguridad.

Si bien la sexta es la medida que repercutirá de forma más clara sobre el sector público, en todo el documento se hace hincapié en el sector público, y prevé un gobierno que preste servicios anticipatorios y personalizados a todos los ciudadanos en todas las etapas de su vida a fin de apoyar el buen funcionamiento de la sociedad. Si se compara con otras estrategias nacionales, el enfoque que ha adoptado Finlandia sitúa la eficiencia del sector público y la eficacia de sus servicios a la par del crecimiento económico (Figura A.2).

[4] http://julkaisut.valtioneuvosto.fi/handle/10024/160391.

Figura A.2: La IA para lograr el correcto funcionamiento de la sociedad

Fuente: www.tekoalyaika.fi/en/reports/finland-leading-the-way-into-the-age-of-artificial-intelligence.

Los objetivos directamente relacionados con la innovación y la transformación del sector público, repartidos en las principales esferas de acción, son los siguientes:

- Desarrollar nuevos modelos operativos para pasar de las actividades basadas en una dependencia a un enfoque que abarque todo el sistema.
- Adaptar el papel del gobierno para asegurar que los ciudadanos tengan el derecho de determinar de forma independiente cómo se utilizan sus datos, al mismo tiempo que se protege la privacidad de los ciudadanos.
- Mejorar la interoperabilidad de los datos gubernamentales y abrir estos datos para impulsar la innovación en todos los sectores; alentar a las empresas a compartir también los datos que posean.
- Crear un Centro de Excelencia para la IA, una universidad virtual de IA y un programa de maestría en IA para fortalecer la reserva de profesionales de primera categoría tanto para el sector privado como para el público.
- Buscar y no detenerse hasta conseguir la creación de una red de asociaciones entre el sector público y el privado para permitir iniciativas de colaboración, intercambio de conocimientos y una mejor adopción del pensamiento multidimensional.
- Celebrar un debate público sobre la ética de la IA en eventos presenciales y en línea.
- Romper los grupos aislados dentro y entre las empresas y los servicios públicos.
- Revisar la legislación en materia de adquisiciones para permitir un desarrollo conjunto efectivo entre el sector público y el privado.

Además, si bien algunas estrategias nacionales se centran en general en estrategias y objetivos, el método de Finlandia también identifica proyectos específicos que debe adoptar el gobierno para facilitar la transformación de la IA, así como los componentes gubernamentales responsables de su implementación. El informe *La Era de la Inteligencia Artificial* [*Age of AI*], de importancia fundamental para el sector público, le pide al gobierno que establezca Aurora, una red de diferentes servicios y aplicaciones inteligentes para "permitirle [a la] administración pública anticiparse mejor y proporcionar recursos para las futuras necesidades de servicio" y dar a los ciudadanos la oportunidad de acceder a servicios digitales de alta calidad las 24 horas del día, los 7 días de la semana.

Expansión hacia el Programa Nacional de Inteligencia Artificial AuroraAI

Desde que se publicó el concepto inicial de Aurora, éste se ha ampliado considerablemente hasta llegar a ser el Programa Nacional de IA AuroraAI. AuroraAI busca ofrecer un conjunto integral de servicios personalizados impulsados por IA para los ciudadanos y las empresas, de una manera que sea humana y trabaje hacia su bienestar como su objetivo final. AuroraAI, visto de manera más general, tiene por objeto permitir a los ciudadanos acceder a la amplia gama de servicios que ofrecen los diversos proveedores de servicios gubernamentales e intersectoriales sin complicaciones. El programa AuroraAI se basa en nueve principios de digitalización (Figura A.3).

Figura A.3: Nueve principios de digitalización



Fuente: https://vm.fi/documents/10623/1464506/AuroraAI+development+and+implementation+plan+2019%E2%80%932023.pdf.

La forma en que los gobiernos tienden a operar, y la forma en que Finlandia operaba en el pasado, es separando las funciones y servicios en dominios distintos, o ministerios, lo que da lugar a enfoques aislados. El programa AuroraAI considera que esto es antitético para un enfoque centrado en el ser humano y para los esfuerzos por mejorar el bienestar integral de sus ciudadanos, ya que el bienestar es multidimensional y, por lo tanto, depende de múltiples dominios. El programa AuroraAI trata de reorientar la prestación de servicios en torno a los ciudadanos y las empresas, combinando datos de múltiples dominios y construyendo una red de aplicaciones de IA centradas en los ciudadanos que presten servicios cuando se necesiten, en torno a diversas actividades empresariales o etapas de la vida y acontecimientos como el nacimiento, la compra de una vivienda o la jubilación.[5] Al reunir datos para crear servicios centrados en el ser humano, "el conocimiento de la situación basado en datos facilita la orientación de servicios eficaces basados en las necesidades reales de las personas y permite a las personas gestionar sus vidas de manera más eficiente en diversas circunstancias de la vida" (AuroraAI, 2019). Esto es posible gracias al uso del aprendizaje por refuerzo (véase el análisis en el Capítulo 2), a través del cual las aplicaciones de la red se perfeccionan con base en la retroalimentación de los usuarios. La red AuroraAI está diseñada para incluir no sólo los servicios públicos sino también los servicios privados y de la sociedad civil.

En la actualidad, AuroraAI se centra en tres eventos de la vida para los programas piloto:

- mudarse a otro sitio para estudiar
- permanecer en el mercado laboral a través de la formación continua
- garantizar el bienestar familiar después de un divorcio.

Cada uno de estos proyectos piloto y sus problemas objetivo asociados, las pruebas piloto y las oportunidades, conclusiones y resultados que se derivan de estos se detallan en el plan de implementación de AuroraAI.[6]

Todo esto se facilita gracias a una personificación digital de los usuarios de AuroraAI llamada DigiMe (Recuadro A.1).

Recuadro A.1: DigiMe

DigiMe se refiere a la capacidad de los ciudadanos de crear un gemelo (o gemelos) digital(es) de sí mismos. Estos personajes digitales les permiten a los usuarios gestionar sus propios datos y utilizarlos para crear perfiles adaptados a la situación con el fin de acceder a servicios personalizados.

La red AuroraAI utiliza el colectivo de estos personajes digitales de forma anónima para identificar similitudes, diferencias y patrones. Luego, estos hallazgos se utilizan para predecir y adaptar de mejor forma los recursos necesarios para prestar servicios anticipados y personalizados a los ciudadanos.

[5] Diríjase a https://youtu.be/IZU_ptEr4eE para ver una presentación sobre AuroraAI por parte del director del programa, Aleksi Kopponen.
[6] https://vm.fi/documents/10623/1464506/AuroraAI+development+and+implementation+plan+2019%E2%80%932023.pdf.

Recuadro A.1: DigiMe *(Cont.)*

Esto se realiza mediante el uso del aprendizaje por refuerzo, a través del cual el sistema identifica qué servicios se necesitan para qué individuos y en qué momentos. Con el paso del tiempo, el sistema recopila información sobre lo que es útil y lo que no y ajusta automáticamente los servicios ofrecidos para que sean más precisos.

Fuente: https://vm.fi/documents/10623/1464506/AuroraAI+development+and+implementation+plan+2019%E2%80%932023.pdf.

Cabe destacar que la estrategia general de Finlandia en materia de IA en relación con los ciudadanos sigue los principios de “MyData”, según los cuales el ciudadano es el único propietario de sus datos. Como propietario, un ciudadano tiene el control total de sus propios datos. Los ciudadanos tienen la facultad para aceptar o rechazar servicios y para tomar decisiones sobre con quién compartir sus datos (AuroraAI, 2019).

En abril de 2019, el gobierno publicó *AuroraAI - Hacia una sociedad centrada en el ser humano* [*AuroraAI – Towards a Human-Centric Society*],[7] que proporciona un plan de implementación de cinco años (2019-23) para AuroraAI. El plan se elaboró en colaboración con una red abierta de más de 330 miembros provenientes de distintos municipios, provincias, organizaciones de la sociedad civil y empresas. A través de él, los autores proponen al próximo gobierno una serie de medidas encaminadas a implementar más plenamente el programa AuroraAI. Entre ellas se incluyen las siguientes:

- Asignar fondos por EUR 100 millones repartidos entre 2020-23 para poner en marcha 10-20 servicios en torno a acontecimientos de la vida y las prácticas comerciales.
- Poner en marcha un proceso de consulta con ciudadanos y empresas para identificar los acontecimientos de la vida y las prácticas empresariales de mayor prioridad, que servirían de base para las actividades de financiamiento y selección de servicios.
- Establecer un equipo de apoyo al cambio y un centro de respuesta central de AuroraAI para apoyar a las dependencias y organizaciones en la implementación de los cambios que lleven al modelo de servicio de AuroraAI.

El plan también prevé un espacio aislado regulatorio para experimentar, de forma controlada, con los datos MyData que hayan sido autorizados por los ciudadanos, así como para explorar la necesidad de realizar algún cambio legislativo para alcanzar el máximo potencial de AuroraAI.

La adopción a corto plazo y el camino a largo plazo de AuroraAI se afirman en varios documentos estratégicos recientes del gobierno. El informe de junio de 2019, titulado *Abriendo camino en la era de la Inteligencia Artificial* [*Leading the Way into the Age of Artificial Intelligence*], compromete al gobierno a trabajar para “garantizar la introducción centrada en el ser humano de la inteligencia artificial y la implementación de principios éticos en el sector público a través del proyecto AuroraAI” durante el próximo año. Además, el Primer Ministro puso en marcha un nuevo programa gubernamental denominado *Finlandia*

[7] https://vm.fi/documents/10623/1464506/AuroraAI+development+and+implementation+plan+2019%E2%80%932023.pdf.

inclusiva y competente [Inclusive and Competent Finland][8] que, entre otras cosas, afirma que "se continuará con el desarrollo seguro y éticamente sostenible de la red AuroraAI, según lo permitan los límites generales de gasto, a fin de facilitar la vida cotidiana y los negocios".

La situación hipotética de "bomba en una caja" de Canadá: Supervisión basada en el riesgo por parte de la IA

Problema

El Ministerio de Transporte de Canadá [*Transport Canada*] es la dependencia responsable de las políticas y programas en materia de transporte del Gobierno del Canadá.[9] Su labor consiste en promover un transporte seguro, eficiente y ambientalmente responsable.

Cada año, el equipo de Identificación de Carga Aérea Previa a la Carga del Ministerio de Transporte de Canadá [*Transport Canada's Pre-load Air Cargo Targeting*] (PACT, por sus siglas en inglés) recibe casi un millón de registros de carga aérea previa a la carga por año, los cuales contienen información como nombre y domicilio del remitente, nombre y domicilio del destinatario, peso y número de piezas. Cada registro puede incluir entre 10 y 100 campos, dependiendo del transportista aéreo y el modelo de negocio del expedidor. Incluso si un empleado trabajara a un ritmo poco realista de un registro por minuto, no tendría ni siquiera tiempo suficiente para revisar el 10 por ciento. Hasta la fecha, muy pocos gobiernos tienen los recursos dedicados a verificar los registros de la carga aérea en busca de riesgos antes de la carga, y de entre aquellos que los tienen, ninguno utiliza la IA. El Ministerio de Transporte de Canadá decidió mejorar esta situación y, de ese modo, aumentar la seguridad del transporte aéreo de carga.[10]

Respuesta

El Ministerio de Transporte de Canadá adoptará a la IA para mejorar los procesos y procedimientos, liberando de estar forma a los empleados para que trabajen en tareas más importantes. La dependencia comenzó explorando el uso de la IA para revisiones basadas en riesgos de los registros de carga aérea, que podrían ampliarse a otras áreas en caso de que fueran exitosas.

Para lograrlo, reunió a un equipo multidisciplinario compuesto por miembros de PACT y de las divisiones de Servicios Digitales y Transformación [*Digital Services and Transformation*] del mismo Ministerio, uno de los Agentes Libres de Canadá[11] y socios de una empresa externa de TI con experiencia en IA. Para el programa piloto, el Ministerio de Transporte de Canadá intentó responder a dos preguntas relacionadas con su desempeño:

---

8 http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161664/Inclusive%20and%20competent%20Finland_2019.pdf.

9 www.tc.gc.ca/en/transport-canada.html.

10 El Ministerio de Transporte de Canadá presentó una descripción de este innovador proyecto a la Plataforma de Estudios de Caso del OPSI (https://oecd-opsi.org/innovations). El contenido de este caso se deriva de su presentación, la cual puede encontrarse en https://oecd-opsi.org/innovations/artificial-intelligence-and-the-bomb-in-a-box-scenario-risk-based-oversight-by-disruptive-technology.

11 El programa de Agentes Libres del Gobierno de Canadá representa un cambio innovador con respecto al modelo de contratación permanente de la administración pública, que organiza el talento y las aptitudes para el trabajo basado en proyectos. Véase el informe de la OPSI en https://oe.cd/innovation2018 para leer estudio de caso completo.

- ¿Puede la IA mejorar nuestra capacidad de llevar a cabo una supervisión basada en riesgos?
- ¿Cómo podemos mejorar la eficacia y la eficiencia al evaluar los riesgos en los envíos de carga aérea?

Para responder a estas preguntas, el equipo de innovación desarrolló e implementó una estrategia de dos etapas en 2018. Como primer paso, utilizaron datos de registros de carga aérea y evaluaciones manuales de riesgos anteriores para explorar enfoques supervisados y no supervisados (véase el Capítulo 2). A través del enfoque supervisado, el equipo intentó comprender la relación entre la información que se ingresaba (registros de carga) y el resultado (es decir, ¿indicaba este registro de carga un mayor nivel de riesgo, dado que se basaba en evaluaciones manuales de riesgo anteriores?) Mediante el aprendizaje no supervisado, el equipo trató de comprender las relaciones entre toda la información que se ingresaba sobre la carga a fin de identificar los envíos raros o inusuales, que podrían ser indicativos de riesgo.

En segundo lugar, el equipo desarrolló una prueba de concepto para poner a prueba el procesamiento del lenguaje natural (PLN) en un subconjunto diferente de datos. El objetivo consistía en poder procesar los registros de carga aérea y etiquetar automáticamente un registro de carga con un indicador de riesgos basado en el contenido de los campos de “texto libre” de los registros de carga aérea y otros campos estructurados. Esto se completó en el primer trimestre de 2018 y demostró que el PLN podía clasificar con éxito los datos de la carga en categorías significativas en tiempo real.

Ambos pasos condujeron a nuevos conocimientos sobre patrones ocultos que pueden indicar riesgos. Como resultado, el equipo pudo usar la IA para generar automáticamente indicadores de riesgo precisos. A través de este programa piloto, el Ministerio de Transporte de Canadá aprendió que la IA era, en efecto, una solución viable para tratar sus cuestiones clave. La dependencia ahora está trabajando para implementar la estrategia en todo su proceso de evaluación de riesgos. Desde la fase de pruebas, el equipo ha generado un tablero y una primera versión de una interfaz de objetivos para identificar cargas potencialmente de alto riesgo.

El equipo se ha asegurado de señalar que la IA no va a reemplazar la actividad humana. La IA se encargará de la categorización, el filtrado y la priorización, lo que actualmente se hace con simples filtros de Excel. La IA es mejor y más eficiente al momento de detectar anomalías, patrones de comercio cambiantes y matices de una manera que una simple hoja de Excel no podría.

El siguiente paso del equipo es realizar pruebas A/B, en las que se comparará la metodología actual con la metodología mejorada por IA. Si tiene éxito, un sistema de selección mejorado por IA podría estar listo ya en marzo de 2021.

Resultados e impacto

Los resultados iniciales fueron muy prometedores. Dado que se puede atender cada uno de los registros de carga, en lugar del pequeño subconjunto posible con las evaluaciones manuales, la IA tiene el potencial de aumentar la seguridad y la protección 15 veces. Además, el PACT puede utilizar la IA para aumentar la capacidad, minimizando al mismo tiempo el número de personas necesarias para hacer el trabajo, haciendo así un mejor uso de los recursos.

Antes de la introducción de la IA, el realizar evaluaciones de riesgos era una tarea onerosa y que requería mucho tiempo. Se necesitaban miles de horas por año para importar, limpiar y archivar datos. Había que disponer de recursos dedicados específicamente al análisis de los registros de carga. Con la introducción de la IA, gran parte de este proceso se ha automatizado y las evaluaciones de riesgos se realizan en tiempo real. La Inteligencia Artificial ayuda a PACT a cumplir con sus resultados de seguridad y hace posible que escaneen más mensajes de carga de más transportistas que nunca antes.

El equipo de innovación considera que este modelo es altamente replicable. Se están llevando a cabo debates preliminares en el seno del Gobierno sobre la adaptación de esta estrategia a otros medios de transporte (p. ej., marítimo, ferroviario, terrestre, etc.) o incluso su ampliación para apoyar el cometido de la dependencia canadiense responsable de las aduanas y la frontera. El equipo dice que, idealmente, todas las oficinas gubernamentales que participan en la seguridad de Canadá, incluidos los organismos de inteligencia, de fronteras y de aplicación de la ley, tendrían acceso a una base de datos única con información que podría utilizarse para optimizar el proceso de supervisión de los envíos de carga en función de los riesgos. Pensando en grande, podría ser algo que se aproveche a nivel internacional.

Desafíos y lecciones aprendidas

Como todos los proyectos de la IA, este programa piloto fue impulsado por los datos. El PACT ya tenía acceso a cantidades importantes de datos; sin embargo, éstos no estaban en un formato que facilitara el uso de la IA. Antes de que la parte de IA del proyecto pudiera comenzar, el equipo tuvo que limpiar y transformar los datos a un formato que pudiera ser consumido por el algoritmo de la IA. Para apoyar la ampliación del proyecto, el equipo está trabajando para hacer frente a este desafío mediante la creación de un canal de datos que alimentará todos los registros de carga recibidos por el Ministerio de Transporte de Canadá en una base de datos única, en un formato que sea inmediatamente consumible por máquina.

Dada la aversión al riesgo en torno a las tecnologías disruptivas en general, también era esencial que el proyecto contara con el apoyo de la alta dirección del Ministerio de Transporte de Canadá. El equipo de innovación encontró ese el apoyo en el Viceministro.

Directrices éticas para una IA fiable de la Comisión Europea

> *Es esencial que la confianza siga siendo el principal cimiento en el que se asienten las sociedades, comunidades y economías, así como el desarrollo sostenible. Por lo tanto, identificamos que la IA fiable constituye nuestra aspiración fundacional, puesto que los seres humanos y las comunidades solamente podrán confiar en el desarrollo tecnológico y en sus aplicaciones si contamos con un marco claro y detallado para garantizar su fiabilidad.*
>
> *Directrices éticas para una IA fiable*

Problema

La Comisión Europea (CE) ha establecido una visión europea para la IA. Uno de sus principales objetivos es aumentar la inversión pública y privada e impulsar su adopción en toda Europa (Comisión Europea, 2018b). La Inteligencia Artificial, especialmente algunos

tipos de Aprendizaje Automático, plantea nuevos tipos de inquietudes sobre ética y equidad en comparación con herramientas anteriores. Es probable que estas inquietudes aumenten a medida que el Aprendizaje Automático se vuelva más universal como resultado de las cantidades cada vez mayores de datos y de la capacidad de procesamiento. La OCDE afirma que uno de los principales retos de la IA es garantizar que los sistemas sean fiables y estén centrados en el ser humano, y ha constatado que se necesitan políticas nacionales para promover sistemas de IA fiables (OCDE, 2019a).

Respuesta

En abril de 2019, la CE publicó las *Directrices Éticas para una IA Fiable*[12] con el fin de proporcionar orientación sobre cómo diseñar e implementar sistemas de IA de manera ética y fiable.

Las *Directrices* fueron creadas por el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (AI HLEG, por sus siglas en inglés) de la CE, que está integrado por 52 expertos en IA procedentes del mundo académico, la sociedad civil y la industria. Una de las tareas fundamentales del AI HLEG ha sido proponer directrices éticas para la IA que tengan en cuenta cuestiones como la equidad, la seguridad, la transparencia, el futuro del trabajo, la democracia, la privacidad y la protección de los datos personales, la dignidad y la no discriminación, entre otras.

Las *Directrices* sostienen que la IA fiable tiene tres componentes que funcionan en armonía:

- **Lícita**. La IA debe cumplir con todas las leyes y reglamentos aplicables.
- **Ética**. La IA debe apegarse a los principios y valores éticos.
- **Robusta**. La IA debe evitar el daño no intencional desde una perspectiva tanto técnica como social.

Además, las *Directrices* se desarrollaron bajo el supuesto de que la ética de la IA se basa en los derechos humanos fundamentales, tal como se establece en los reglamentos de la UE y en el derecho internacional de los derechos humanos. En consecuencia, el proceso puso de manifiesto cuatro principios éticos que deben tenerse en cuenta al diseñar y desplegar la IA (véase el recuadro A.2).

Recuadro A.2: Principios éticos e imperativos identificados por las *Directrices*

1. Respeto de la autonomía humana

Los humanos que interactúan con los sistemas de IA deben ser capaces de mantener una autonomía plena y efectiva sobre sí mismos. Los sistemas de IA no deberían subordinar, coaccionar, engañar, manipular, condicionar o dirigir a los seres humanos de forma injustificada. En cambio, deberían aumentar, complementar y potenciar las aptitudes humanas. La asignación de funciones entre los seres humanos y los sistemas de IA debería seguir principios de diseño centrados en el ser humano y dejar una oportunidad significativa para la elección humana. Esto significa garantizar la

[12] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

Recuadro A.2: Principios éticos e imperativos identificados por las *Directrices* *(Cont.)*

supervisión humana de los sistemas de IA. Debería apoyar a los seres humanos en el entorno laboral y aspirar a la creación de trabajos significativos.

2. Prevención del daño

Los sistemas de IA no deberían causar o exacerbar daños ni perjudicar a los seres humanos de ninguna otra manera. Esto implica proteger la dignidad del ser humano, así como su integridad física y mental. Los sistemas de IA y sus entornos deben ser seguros y estar protegidos. Deben ser robustos desde el punto de vista técnico y no estar abiertos a usos malintencionados. Las personas vulnerables deberían recibir mayor atención y participar en el desarrollo, despliegue y uso de la IA. Debe prestarse especial atención a los efectos adversos causados por las asimetrías de poder o de información, como las que existen entre los gobiernos y los ciudadanos.

3. Equidad

El desarrollo, despliegue y uso de los sistemas de IA deben ser justos. Sustancialmente, esto implica el compromiso de 1) garantizar una distribución equitativa y justa tanto de los beneficios como de los costos, y 2) garantizar que los individuos y grupos estén libres de prejuicios injustos, discriminación y estigmatización. Si se pueden evitar los prejuicios injustos, los sistemas de IA podrían incluso mejorar la justicia social. También debe fomentarse la igualdad de oportunidades en términos de acceso a la educación, los bienes, los servicios y la tecnología. El uso de la IA nunca debe ocasionar que se engañe o perjudique a las personas de forma injustificada en cuanto a su libertad de elección. La equidad implica que los profesionales de la IA deben respetar el principio de proporcionalidad entre los medios y los fines, además de considerar cuidadosamente cómo equilibrar los intereses y objetivos en conflicto. Desde el punto de vista del procedimiento, la equidad conlleva la capacidad de apelar las decisiones tomadas por los sistemas de IA y los humanos que los operan.

4. Explicabilidad

Los procesos deben ser transparentes, las capacidades y el propósito de los sistemas de IA deben comunicarse abiertamente, y las decisiones resultantes deben poder explicarse a los afectados, en la medida de lo posible. De lo contrario, no es posible impugnar una decisión. Sin embargo, no siempre es posible explicar por qué y cómo un modelo ha generado una decisión en particular. Estos casos se denominan algoritmos de “caja negra” y requieren una atención especial. En esas circunstancias, pueden requerirse otras medidas de explicabilidad (por ejemplo, rastreabilidad, auditabilidad y comunicación transparente sobre las capacidades del sistema), siempre que el sistema en su conjunto respete los derechos fundamentales. El grado en que se necesita la explicabilidad depende del contexto y de las consecuencias en caso de que el resultado sea inexacto.

Fuente: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477 (según el extracto de la OCDE).

Para orientar a las organizaciones interesadas en el uso de la IA, las *Directrices* establecen siete requisitos que los sistemas de IA deben cumplir para ser considerados fiables (Recuadro A.3). Estos están diseñados para ayudar a las organizaciones a materializar los cuatro principios clave.

Recuadro A.3: Siete requisitos para una IA fiable

1. **Acción y supervisión humanas**
   - Si un sistema de IA tiene el potencial de afectar de forma negativa los derechos humanos, se debe evaluar su impacto sobre los derechos fundamentales antes de su desarrollo.
   - Los seres humanos deben ser capaces de tomar decisiones autónomas informadas sobre los sistemas de IA y, cuando sea necesario, desafiarlos. Asimismo, las personas deben tener el derecho a no ser objeto de una decisión exclusivamente automatizada si ésta les afecta de manera significativa.
   - Los seres humanos deben tener la capacidad de supervisar (en grados variables según el área de aplicación y el riesgo) los sistemas de IA.

2. **Solidez técnica y seguridad**
   - Los sistemas de IA deben desarrollarse de manera que se busque prevenir los riesgos, promover un comportamiento fiable y minimizar o prevenir daños.
   - Se deben establecer procesos de seguridad para proteger los sistemas de IA contra las vulnerabilidades que puedan ser objeto de abuso (por ejemplo, hackeos) y para evitar usos no deseados del sistema.
   - Se deben establecer procesos para evaluar el posible riesgo, y los sistemas de IA deben contar con salvaguardas en caso de problemas (por ejemplo, requisitos para la intervención humana en algunas circunstancias).
   - Se debe establecer un proceso explícito para atender los riesgos no previstos derivados de resultados y predicciones inexactos.
   - Los resultados del sistema de IA deben ser reproducibles y fiables.

3. **Gestión de la privacidad y de los datos**
   - Los sistemas de IA deben garantizar la privacidad y la protección de datos a lo largo de su ciclo de vida para evitar la discriminación ilegal o injusta.
   - Deben hacerse esfuerzos por garantizar la calidad de los datos y corregir cualquier sesgo, imprecisión o error. La integridad de los datos también debe garantizarse para brindar protección, por ejemplo, contra la introducción de datos maliciosos en un sistema.
   - Las organizaciones deben contar con protocolos de gestión de datos que rijan el acceso a los mismos.

4. **Transparencia**
   - Los conjuntos de datos y los procesos de toma de decisiones deberían documentarse en la medida de lo posible para permitir la rastreabilidad y la transparencia.
   - En aquellos casos en que los sistemas de IA tengan un impacto en la vida de las personas, los afectados deberían poder exigir una explicación del proceso de toma de decisiones.

Recuadro A.3: Siete requisitos para una IA fiable *(Cont.)*

- Las personas tienen derecho a saber que están interactuando con un sistema de IA, así como a estar informados sobre las capacidades y limitaciones del mismo.

5. Diversidad, no discriminación y equidad

- Se deben establecer procesos y procedimientos para atender y eliminar los sesgos en la fase de recopilación de datos cuando sea posible, así como mecanismos de supervisión para monitorear los sistemas.
- Según la finalidad del sistema de IA, todos los usuarios deberían tener un acceso equitativo a los productos de IA, independientemente de sus datos demográficos o características.
- Las organizaciones deben consultar a las partes interesadas que puedan verse afectadas por un sistema de IA a lo largo de su ciclo de vida para obtener retroalimentación periódica.

6. Bienestar social y ambiental

- Las organizaciones deben tomar decisiones sostenibles para los sistemas de IA (por ejemplo, el consumo de energía), teniendo en cuenta su ciclo de vida completo y la cadena de suministro.
- Se deben monitorear y considerar los efectos sociales de los sistemas de IA (por ejemplo, intervención social, relaciones sociales).
- Se deben considerar los efectos sociales y democráticos de los sistemas de IA (por ejemplo, su efecto sobre las instituciones, la democracia y la sociedad).

7. Rendición de cuentas

- Debe ser posible evaluar los algoritmos, los datos y los procesos de diseño.
- Las organizaciones deben tratar de identificar, evaluar, documentar y minimizar los posibles efectos perjudiciales de los sistemas de IA.
- Deben establecerse métodos para negociar, evaluar y documentar los casos en que surjan tensiones entre estos requisitos, y en los que tal vez sea necesario hacer concesiones. Si no es posible encontrar un equilibrio ético aceptable, el sistema de IA no deberá continuar de esa forma.
- Deben establecerse mecanismos que garanticen a las personas el derecho a la reparación de daños cuando se produzca un impacto adverso injusto.

Las *Directrices* recomiendan que estos requisitos sean evaluados y atendidos continuamente a lo largo del ciclo de vida de un sistema de IA. Para ayudar a las organizaciones a cumplir con estos requisitos, las *Directrices* describen los métodos técnicos (por ejemplo, la arquitectura de los sistemas, los métodos de explicación) y los métodos no técnicos (por ejemplo, la participación de las partes interesadas, los códigos de conducta) necesarios para lograrlos.

Fuente: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477 (adaptado por la OCDE).

Por último, las *Directrices* proporcionan una "lista de evaluación" concreta diseñada para ayudar a los desarrolladores de sistemas de IA orientados al usuario, a poner en práctica los requisitos clave expuestos en el Recuadro A.4. Esta lista se encuentra actualmente en un proceso piloto digital[13] para su prueba y validación. Se invita a todos los interesados en la lista a que proporcionen información sobre las formas en que puede reforzarse. La CE tiene previsto evaluar toda la retroalimentación recibida hasta finales de 2019 para incorporarla a una nueva versión en 2020. En el recuadro A.4 figuran extractos de la lista de evaluación. Se invita a los lectores de este manual a que lean y consideren la lista completa en el documento de las *Directrices*.[14]

Recuadro A.4: Ejemplos seleccionados de la Lista para la evaluación de la fiabilidad de la IA (versión piloto)

1. Acción y supervisión humanas

- ¿Ha considerado usted si el sistema de IA debería informar a los usuarios que una decisión o resultado es fruto de una decisión algorítmica?
- ¿Ha tenido usted en cuenta la asignación de tareas entre el sistema de IA y los humanos para garantizar interacciones adecuadas?
- ¿Se ha asegurado de disponer de un botón de desconexión o un procedimiento que permita abortar una operación en caso necesario? ¿Este procedimiento aborta el proceso o delega el control a un humano?

2. Solidez técnica y seguridad

- ¿Ha adoptado medidas o sistemas para garantizar la integridad del sistema de IA y su capacidad para resistir posibles ataques?
- ¿Se ha asegurado de que su sistema cuente con un plan de repliegue suficiente en caso de que se enfrente a algún ataque malintencionado u otro tipo de situación de este tipo (por ejemplo, procedimientos técnicos de conmutación o la solicitud de un operador humano antes de continuar)?
- ¿Ha verificado el daño que se causaría si el sistema de IA realizara predicciones inexactas?
- ¿Ha comprobado si es necesario tener en cuenta algún contexto o condición en particular para garantizar la reproducibilidad?

3. Gestión de la privacidad y de los datos

- ¿Ha analizado formas de desarrollar el sistema de IA o de entrenar el modelo con un uso mínimo de datos potencialmente sensibles?
- ¿Ha alineado su sistema con las normas pertinentes (p. ej., ISO, IEEE) o con los protocolos generalmente adoptados para la gestión y gobernanza cotidianas de los datos?
- ¿El sistema registra cuándo, dónde, cómo y quién accede a los datos, así como con qué propósito?

[13] https://ec.europa.eu/futurium/en/ethics-guidelines-trustworthy-ai/register-piloting-process-0.
[14] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477.

Recuadro A.4: Ejemplos seleccionados de la Lista para la evaluación de la fiabilidad de la IA (versión piloto) *(Cont.)*

4. Transparencia

- ¿Ha adoptado medidas que puedan garantizar la rastreabilidad, como la documentación de los métodos de entrenamiento del algoritmo?
- ¿Ha evaluado en qué medida se pueden comprender las decisiones y, por tanto, los resultados del sistema de IA?
- ¿Ha comunicado con claridad las características, limitaciones y posibles deficiencias del sistema de IA?

5. Diversidad, no discriminación y equidad

- ¿Ha tenido en cuenta la diversidad y representatividad de los usuarios en los datos? ¿Ha realizado pruebas para poblaciones específicas o casos de uso problemático?
- ¿Ha evaluado si el equipo que participó en la construcción del sistema de IA es representativo de la audiencia a la que está dirigido? ¿Es representativo de la población en general, teniendo en cuenta otros a grupos que podrían verse afectados de manera tangencial?
- ¿Ha estudiado la posibilidad de introducir algún mecanismo para incluir la participación de diferentes partes interesadas en el desarrollo y el uso del sistema de IA?

6. Bienestar social y ambiental

- ¿Se ha asegurado de introducir medidas para reducir el impacto ambiental del ciclo de vida de su sistema de IA?
- ¿Se ha asegurado de que el sistema de IA indique claramente que su interacción social es simulada y que no tiene capacidad de “entender” ni “sentir”?
- ¿Ha evaluado el impacto social más amplio asociado al uso del sistema de IA más allá del que tenga sobre el usuario individual (por ejemplo, las partes interesadas pueden verse afectadas de forma indirecta)?

7. Responsabilidad

- En los casos en que el uso del sistema afecta a los derechos fundamentales, ¿se aseguró de que el sistema de IA pueda ser auditado de manera independiente?
- ¿Existen procesos para que terceros o los trabajadores informen sobre posibles vulnerabilidades, riesgos o sesgos?
- ¿Cómo toma las decisiones sobre las concesiones necesarias? ¿Se ha asegurado de que se documentara la decisión sobre tales concesiones?
- ¿Ha establecido un conjunto adecuado de mecanismos que permita la reparación de daños en el caso de que se produzca cualquier daño o efecto adverso?

Fuente: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/2 (adaptado por la OCDE).

Aunque son bastante detalladas, los autores del AI HLEG toman la precaución de señalar que será necesario adaptar las *Directrices* a cada situación específica. Esto es importante, por ejemplo, porque algunas aplicaciones de IA son más sensibles que otras.

Las *Directrices* se elaboraron a través de un proceso abierto y participativo. En diciembre de 2018 se publicó un borrador inicial para su consulta pública, y se recibieron más de 500 comentarios provenientes de un conjunto diverso de participantes, desde empresas y organizaciones de la sociedad civil hasta miembros del público en general.[15] Estos comentarios se tuvieron en cuenta en la creación del documento final de las *Directrices*.

Directiva sobre la toma de decisiones automatizada [*Directive on Automated Decision-Making*] de Canadá

Estas son sus barandillas para una automatización responsable[16]

Alex Benay, Exdirector de Sistemas de Información del Gobierno de Canadá (2007-2011)

Problema

El Gobierno de Canadá (GC) está tratando de utilizar cada vez más la Inteligencia Artificial para tomar o ayudar en la toma de decisiones administrativas a fin de mejorar la prestación de servicios.[17] Sin embargo, una cuestión clave en este sentido es la posibilidad de que existan sesgos, problemas de ética y otras cuestiones a medida que los organismos gubernamentales avanzan en la adopción de la IA, así como cuestiones sobre el grado en que los humanos deberían participar en la toma de decisiones basada en la IA. Las leyes y políticas existentes pueden complementarse con orientación adicional para hacer frente a los desafíos específicos de la tecnología y los casos de uso.

Estos desafíos surgieron y se consideraron cuidadosamente en el lanzamiento de proyectos piloto gubernamentales para desarrollar algoritmos avanzados que ayudaran a priorizar las solicitudes de visado de residente temporal de China y la India. El número de solicitudes había aumentado considerablemente en esas regiones, lo que suponía una carga para la tramitación. Los proyectos produjeron resultados y lecciones valiosas para el gobierno, entre otras cosas, al proporcionar una base de orientación normativa para ayudar a garantizar la implementación responsable del apoyo a la toma de decisiones basado en la IA en el sistema de inmigración de Canadá. Dado que otras esferas del Gobierno de Canadá recurren a la IA para mejorar la eficiencia y el servicio al cliente, se necesitaba una orientación más amplia para ayudar a las dependencias a identificar y atender las cuestiones de parcialidad, transparencia, entre otras, en sus propios contextos.

En términos más generales, las empresas con mayor experiencia tecnológica, como Google, también han tenido problemas para implementar soluciones de IA.[18] Esto demuestra además la necesidad de establecer salvaguardas claras para garantizar que la IA y las tecnologías basadas en datos, no den lugar a resultados indeseables.

---

[15] Para obtener más información sobre la consulta pública, incluido un resumen de la retroalimentación recibida, véase https://ec.europa.eu/digital-single-market/en/news/over-500-comments-received-draft-ethical-guidelinestrustworthy-artificial-intelligence.

[16] https://askai.org/blog-podcast-canada-cio-alex-benay-is-on-a-mission-to-modernize.

[17] www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

[18] www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas.

Con el crecimiento de los servicios habilitados por IA, incluidos aquellos que toman decisiones automáticas que afectan la vida de seres humanos, el gobierno está tratando de garantizar que dichas decisiones minimicen los riesgos para los ciudadanos, así como para los organismos del sector público a nivel federal. El objetivo es proporcionar una base ética clara para los organismos gubernamentales en los casos relacionados con la toma de decisiones automatizada, con el fin de prevenir situaciones como las anteriores.

Respuesta

El Gobierno de Canadá recurrió a la investigación de cientos de expertos en informática y funcionarios gubernamentales para elaborar el libro blanco *Inteligencia artificial responsable en el Gobierno de Canadá* [*Responsible Artificial Intelligence in the Government of Canada*],[19], y el 1º de abril de 2019 publicó su Directiva sobre la toma de decisiones automatizada.[20] La Directiva proporciona un enfoque basado en riesgos para garantizar la transparencia, la rendición de cuentas, la legalidad y la equidad de las decisiones automatizadas que afectan a los canadienses, e impone ciertos requisitos para el uso de algoritmos y sistemas de toma de decisiones por parte del gobierno. La Directiva es la primera de este tipo en el mundo, y entrará en vigor en todo el Gobierno Federal a partir de abril de 2020.

Actualmente, la postura inicial del gobierno en cuanto a sus políticas incluye sistemas automatizados de toma de decisiones centrados en el público, un área del Gobierno de Canadá que tenía una gran necesidad de una gestión basada en riesgo. Entre los ejemplos figuran los programas de prestaciones que deciden si los solicitantes cumplen los requisitos para beneficiarse de las mismas. Las oportunidades y los problemas futuros que se introduzcan por el uso de la Inteligencia Artificial en el sector público se abordarán de manera iterativa a medida que evolucionen las necesidades.

El pilar de la Directiva es la Evaluación de Impacto Algorítmico (véase el Recuadro A.5), que los dirigentes de las dependencias deben completar antes de producir o modificar de forma significativa, un sistema automatizado de toma de decisiones. Esto les ayuda a adelantarse a los problemas y a establecer sistemas y procesos para monitorear la implementación. Los dirigentes deben publicar los resultados de sus evaluaciones en línea como datos abiertos del gobierno. Los funcionarios del GC destacan la importancia de este componente, el cual proporciona una visión general completa y pública de los programas automatizados de toma de decisiones existentes.

Recuadro A.5: Evaluación de Impacto Algorítmico

Esta Evaluación es un cuestionario digital que evalúa el posible impacto de un sistema automatizado de decisión centrado en el público. Ésta evalúa las decisiones que el sistema tiene la capacidad de tomar o informar y el daño potencial a los ciudadanos. Los resultados del cuestionario generan una calificación del impacto en una escala de 1 a 4 con respecto al sistema de toma de decisiones: 1 indica decisiones que producen un impacto breve y reversible, mientras que 4 indica decisiones que podrían tener un impacto irreversible y significativo. Esta calificación establece el nivel mínimo

[19] https://docs.google.com/document/d/1Sn-qBZUXEUG4dVk909eSg5qvfbpNlRhzIefWPtBwbxY.
[20] www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

Recuadro A.5: Evaluación de Impacto Algorítmico *(Cont.)*

de responsabilidad de la dependencia y asigna requisitos obligatorios de gestión, supervisión y presentación de informes.

Las preguntas de muestra incluyen:

- ¿Está el proyecto dentro de un área sometida a una constante supervisión rigurosa por parte del público?
- ¿Son los clientes de esta línea de negocios particularmente vulnerables?
- ¿Será difícil de interpretar o explicar el proceso algorítmico?
- ¿Reemplazará el sistema a una decisión que de otra manera sería tomada por un humano?
- ¿Son reversibles los impactos resultantes de la decisión?
- ¿Quién recopiló los datos utilizados para entrenar el sistema?
- ¿Contará usted con procesos documentados para poner a prueba los conjuntos de datos con el fin de detectar sesgos y otros resultados inesperados?
- ¿Proporcionará el sistema una pista de auditoría que registre todas las recomendaciones o decisiones que éste mismo haya hecho o tomado?
- ¿Podrá el sistema presentar las razones de sus decisiones o recomendaciones cuando así se solicite?
- ¿Permitirá el sistema que los humanos anulen sus decisiones?

Con base en las respuestas a estas y otras preguntas, la evaluación especifica la respuesta requerida en función del impacto potencial. Por ejemplo, determina la medida en que se necesita:

- una revisión por pares del sistema
- un aviso público sobre el sistema
- la participación humana en el proceso de toma de decisiones
- una explicación de cómo se toman las decisiones
- someter el sistema a pruebas y vigilar los resultados para detectar resultados inesperados (p. ej., sesgos)
- capacitar al personal para que entienda y supervise el sistema
- establecer planes de contingencia
- aprobación para operar.

El uso de la herramienta de evaluación será obligatorio en Canadá a partir de abril de 2020.

La herramienta está disponible en el Portal de Gobierno Abierto [*Open Government Portal*] y se puede encontrar como software libre y de código abierto (FOSS, por sus siglas en inglés) en GitHub. El Gobierno de Canadá alienta a los gobiernos de otros países, a expertos y a grupos comunitarios a que participen en el desarrollo y la evolución continuos de la herramienta, y los invita a que la adapten para ajustarla a sus propios contextos institucionales y culturales.

Fuente: www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592; https://canada-ca.github.io/aia-eia-js; https://open.canada.ca/data/en/dataset/748a97fb-6714-41ef-9fb8-637a0b8e0da1; entrevistas con funcionarios del GC.

Además, la Directiva exige a los organismos gubernamentales que liberen el código fuente personalizado de los algoritmos en la medida de lo posible, y que proporcionen a los clientes las opciones de recurso aplicables para apelar las decisiones, entre otras cosas.

Cabe destacar que la Directiva y la Herramienta de Evaluación se elaboraron de manera abierta y participativa. Se invitó a las partes interesadas de todos los sectores, así como a los miembros del público, a formular observaciones. Esto permitió incorporar durante el proceso de desarrollo la retroalimentación de los círculos académicos, las organizaciones de la sociedad civil, las empresas del sector privado y las personas interesadas (Gobierno de Canadá, 2019a).

Resultados e impacto

Aunque la Directiva entrará plenamente en vigor hasta abril de 2020, los organismos gubernamentales ya han cambiado sus conductas para garantizar su cumplimiento. Esto incluye completar la Evaluación de Impacto Algorítmico y seguir los requisitos basados en el riesgo establecidos conforme a la clasificación de riesgo determinada por la Evaluación.

El impacto de esta estrategia de Canadá ya se está difundiendo a nivel internacional. Todos los miembros de D9[21], una red de las naciones digitales más avanzadas del mundo, están considerando la posibilidad de adoptar la Evaluación de Impacto Algorítmico (Greenwood, 2019). Los Gobiernos de Alemania y México han adoptado versiones modificadas de la Evaluación de Impacto Algorítmico adaptadas a sus propios contextos.

Desafíos y lecciones aprendidas

El equipo del GC que diseñó la Directiva y la Evaluación de Impacto Algorítmico no se enfrentó a muchos retos significativos durante las etapas de desarrollo y aprobación. Atribuyen esto a unos pocos factores clave en su estrategia:[22]

- **Intervención oportuna**: El equipo reconoció desde etapas tempranas, la falta de políticas y actuó rápidamente para desarrollar los instrumentos necesarios en materia de políticas. Al centrarse en un área problemática precisa y bien definida, fueron capaces de obtener los resultados deseados con rapidez antes de una implementación importante. Esto facilitó la elaboración de la Directiva y garantizó que existiera la orientación necesaria cuando el campo de la IA comenzó a generar mayor interés.
- **Trabajo abierto:** Según los funcionarios del GC, la apertura fue fundamental para lograr una amplia aceptación de la Directiva, así como de las clasificaciones basadas en el riesgo determinadas por la Evaluación de Impacto Algorítmico. Este proceso abierto también aceleró el desarrollo de la Directiva y la herramienta de Evaluación al aprovechar la experiencia de los colaboradores externos.
- **Hacer antes de mostrar:** Para garantizar que se contara con apoyo político, el equipo gubernamental creó un primer borrador de la Directiva y un prototipo de la Evaluación de Impacto Algorítmico, y los presentó a los altos dirigentes. Este enfoque funcionó mejor que tratar de describir el concepto y la necesidad asociada en una propuesta.

[21] Véase www.digital.govt.nz/digital-government/international-partnerships/the-digital-9.
[22] Entrevista con Michael Karlin, Jefe de Equipo, Política de Datos del Departamento de Defensa Nacional, Gobierno del Canadá, 5 de junio de 2019.

A partir de la experiencia del gobierno en el desarrollo de esta Directiva también se obtuvieron aprendizajes más generales:

- **Guiar y facultar en lugar de limitar y restringir:** Si bien los organismos públicos están obligados a cumplir con la Directiva, en sí ésta no les impide emprender (o restringir) el desarrollo de su propio sistema automatizado de toma de decisiones. No obstante, la Directiva faculta a las autoridades y a terceros (del sector privado o de la sociedad civil) para plantear preguntas importantes sobre el sistema propuesto que los desarrolladores deben responder (por ejemplo, qué datos se utilizan, cómo se utilizan y si el sistema cumple con la normativa vigente).
- **La IA no es la fase final:** Aunque la llegada de la Inteligencia Artificial y sus desafíos fueron lo que impulsó la creación de la Directiva, la iniciativa debe considerarse como parte de un esfuerzo más amplio del gobierno por proteger a sus ciudadanos y al mismo tiempo, promover la innovación para mejorar los servicios públicos.
- **Centrarse en las consecuencias de la automatización y no en la tecnología:** Al redactar la Directiva, el equipo se centró en la pregunta "¿Qué sucede cuando delegamos las decisiones humanas a una máquina?" más que en la propia tecnología de IA. Si bien algunas áreas, como la necesidad de comprobar la existencia de sesgos en los datos de aprendizaje, están informadas por los riesgos de la IA, la Directiva evita referirse a tecnologías específicas. Esto ayuda a asegurar una amplia aplicabilidad, desde los sistemas simples basados en reglas hasta los más complejos sistemas de aprendizaje profundo, así como una relevancia a largo plazo, a medida que la tecnología evoluciona.
- **Construir un ecosistema, no sólo una política:** De forma paralela al desarrollo de la política, el Ministerio de Servicios Públicos y Adquisiciones de Canadá [*Public Service and Procurement Canada*] (PSPC, por sus siglas en inglés), junto con la Secretaría del Consejo del Tesoro de Canadá [*Treasury Board of Canada Secretariat*] (TBS, por sus siglas en inglés), llevó a cabo un proceso de adquisiciones para establecer una lista de proveedores que pudieran proporcionar al Gobierno de Canadá servicios, soluciones y productos de IA responsables y eficaces. Los proveedores se comprometieron a apoyar los esfuerzos del Gobierno de Canadá por abrir el camino de la IA ética.

Además, la Academia Digital, una dependencia de enseñanza con sede en la Escuela de Administración Pública de Canadá, ofreció un nuevo plan de estudios centrado en los datos, el diseño, el desarrollo, la Inteligencia Artificial y el Aprendizaje Automático, las DevOps y la tecnología disruptiva. Este plan de estudios es un elemento importante de la siguiente fase de renovación del servicio público, "Más allá de 2020" [*Beyond 2020*], que se centra en la importancia de un servicio público ágil, inclusivo y equipado.

Estrategia y hoja de ruta del gobierno de los Estados Unidos en materia de datos

> *La misión de la Estrategia Federal de Datos es aprovechar todo el valor de los datos federales para cumplir la misión, y por el bien del servicio y el bien público, guiando al Gobierno Federal en la práctica de la gobernanza ética, el diseño consciente y una cultura de aprendizaje.*
>
> *Estrategia Federal de Datos de los Estados Unidos*

Problema

El Gobierno de los Estados Unidos es una de las entidades más grandes del mundo, lo que puede dificultar la gestión de los datos como un activo de manera uniforme a nivel empresarial. Recientemente se han aprobado varias nuevas leyes que tratan de atender esta situación y que podrían complementarse con una orientación normativa uniforme en el poder ejecutivo, para ayudar a garantizar una implementación coherente. Además de las nuevas leyes, el presidente de los Estados Unidos ha identificado el hecho de "aprovechar los datos como un activo estratégico" como un área de prioridad presidencial y un Objetivo Prioritario Interinstitucional (Objetivo CAP, por sus siglas en inglés) que requiere un enfoque de sistemas para los datos en el gobierno.[23]

Respuesta

El 4 de junio de 2019, la Oficina de Administración y Presupuesto de la Casa Blanca [*White House Office of Management and Budget*] (OMB, por sus siglas en inglés) puso en marcha la Estrategia de Datos Federales (la Estrategia) como marco de todo el gobierno para ayudar a promover la coherencia y la calidad de la infraestructura, la gestión, las acciones, la protección y la seguridad de datos. La Estrategia fue creada por un equipo intergubernamental[24] y representa una visión a diez años de cómo el gobierno "acelerará el uso de los datos para apoyar los fundamentos de la democracia, cumplir con su misión, servir al público y administrar los recursos, protegiendo al mismo tiempo la seguridad, la privacidad y la confidencialidad".

La Estrategia consta de **10 principios** organizados en torno a tres categorías que sirven como directrices de motivación para los organismos gubernamentales (véase el recuadro A.6).

Recuadro A.6: Principios de la Estrategia

Gobierno ético

1. **Mantener la ética**: Monitorear y evaluar las repercusiones sobre el público que tendrá la utilización de los datos. Diseñar controles y contrapesos para proteger y servir al bien público.
2. **Ejercer la responsabilidad:** Practicar una administración y gestión eficaces. Garantizar la seguridad, la privacidad, la confidencialidad prometida y las prácticas apropiadas de acceso y uso.
3. **Promover la transparencia**: Expresar el propósito y los usos que se le dan a todos los datos para generar confianza pública. Documentar todos los procesos y productos para informar a los proveedores y usuarios de los datos.

[23] www.performance.gov/CAP/leveragingdata. Los Objetivos CAP son metas a largo plazo utilizadas por los dirigentes para acelerar el progreso en un número limitado de esferas prioritarias de la Presidencia, en las que su aplicación requiere la colaboración activa de múltiples organismos. Tratan de impulsar la colaboración entre los gobiernos para hacer frente a los problemas de gestión de todo el gobierno.
[24] https://strategy.data.gov/team.

Recuadro A.6: Principios de la Estrategia *(Cont.)*

Diseño consciente

4. **Asegurar la relevancia:** Proteger la calidad y la integridad de los datos. Asegurar que los datos sean adecuados, precisos, objetivos, útiles, comprensibles y oportunos.
5. **Aprovechar los datos existentes:** Identificar las necesidades para informar las cuestiones de investigación y de políticas, reutilizando los datos si es posible.
6. **Anticipar los usos futuros**: Considerar y planificar la reutilización y la interoperabilidad desde el principio.
7. **Demostrar capacidad de respuesta:** Mejorar los datos a través de aportaciones de los usuarios. Mediante un proceso cíclico de retroalimentación, establecer una línea de base, obtener apoyo, colaborar y perfeccionar de manera continua.

Cultura de aprendizaje

8. **Invertir en el aprendizaje:** Promover una cultura de aprendizaje continuo y colaborativo mediante la inversión continua en infraestructura y recursos humanos.
9. **Desarrollar líderes de datos**: Cultivar el liderazgo a todos los niveles invirtiendo en la capacitación y el desarrollo del valor de los datos para las misiones, el servicio y el bien público.
10. **Practicar la responsabilidad**: Asignar responsabilidades, auditar las prácticas de datos, documentar y aprender de los resultados, y hacer los cambios necesarios.

Fuente: www.whitehouse.gov/wp-content/uploads/2019/06/M-19-18.pdf (editado para brevedad).

Junto con los principios, la Estrategia elaboró 40 prácticas para orientar a las dependencias sobre cómo aprovechar el valor de los datos federales, así como de los datos patrocinados por el gobierno federal. Las prácticas toman en cuenta los diferentes usos de los datos disponibles para lograr un mejor valor público. Con ello, las prácticas buscan armonizar la gestión de los datos con esos usos a fin de atender a las necesidades tanto del gobierno como de las partes interesadas/usuarios. Las prácticas se agrupan en torno a tres categorías básicas. En el Recuadro A.7 se enumeran esas categorías y se presenta una muestra de las prácticas establecidas en la Estrategia. Las 40 prácticas figuran en la Estrategia completa.

Recuadro A.7: Prácticas seleccionadas de la Estrategia

Construir una cultura que valore los datos y promueva el uso público: Esta práctica se centra en la utilización de los datos para la toma de decisiones gubernamentales y en promover el uso externo.

- **Identificar las necesidades de datos para responder a las preguntas clave de las dependencias:** Identificar y priorizar las preguntas clave y los datos necesarios para responderlas.

Recuadro A.7: Prácticas seleccionadas de la Estrategia *(Cont.)*

- **Monitorear y atender la percepción pública:** Evaluar regularmente la confianza del público en términos de valor, precisión, objetividad y protección de la privacidad para hacer mejoras y avanzar en el cumplimiento de las misiones.
- Conectar las funciones de datos entre dependencias: **Establecer comunidades de práctica para funciones comunes** (por ejemplo, gestión de datos, análisis) y para promover la eficiencia, la colaboración y la coordinación.

Administrar, gestionar y proteger los datos: Esta práctica se centra en la gestión de los datos en todas las dependencias.

- **Priorizar la gestión de los datos:** Garantizar la eficiencia de las autoridades, los puestos, las estructuras, las políticas y los recursos para apoyar de manera transparente la gestión, el mantenimiento y la utilización de los datos.
- **Permitir la modificación:** Establecer procedimientos claros que permitan a los miembros del público acceder a los datos sobre ellos mismos y modificarlos, según corresponda.
- **Compartir datos entre los gobiernos estatales, locales y tribales, y las dependencias federales:** Facilitar el intercambio de datos, particularmente en el caso de los programas financiados por el gobierno federal y administrados localmente.

Promover el uso eficiente y apropiado de los datos: Esta práctica se centra en dar acceso a los recursos de datos (p. ej., intercambio de datos, datos abiertos), promover su uso apropiado (documentar y proteger los datos) y brindar orientación sobre el aumento de datos (calidad de los datos, metadatos, vínculos seguros).

- **Aumentar la capacidad de gestión y análisis de datos:** Educar a la fuerza de trabajo mediante capacitación, herramientas, comunidades y la ampliación de sus capacidades.
- **Alinear la calidad con el uso previsto:** Asegurar que los datos que puedan servir de base para decisiones importantes sobre políticas o del sector privado, sean de utilidad, integridad y objetividad adecuadas.
- **Diversificar los métodos de acceso a los datos:** Invertir en múltiples niveles de acceso para que los datos sean lo más accesibles posible.

Fuente: www.whitehouse.gov/wp-content/uploads/2018/10/M-19-01.pdf.

Para que estas prácticas sean aplicables, la Estrategia requiere que las dependencias se apeguen a los requisitos de los planes de acción anuales de todo el gobierno, en los que se da prioridad a las prácticas para un año determinado y se establecen plazos y se asignan responsabilidades. Junto con la publicación de la Estrategia, Estados Unidos publicó una versión preliminar del primero de estos planes, el *Plan de Acción de la Estrategia de Datos Federales 2019-2020* [*2019-2020 Federal Data Strategy Action Plan*].[25] En el proyecto, que comprendió un periodo de consulta pública y de las partes interesadas de tres semanas de duración, hasta el 8 de julio de 2019, se enumeró una serie de 16 acciones concretas en

[25] El proyecto de plan está disponible en https://strategy.data.gov/assets/docs/draft-2019-2020-federal-data-strategy-action-plan.pdf. Se espera que la versión final del plan se publique en https://strategy.data.gov/action-plan.

todo el gobierno, incluyendo un enfoque en la Inteligencia Artificial. Consiste en tres tipos de acciones:

1. **Compartidas** - dirigidas por una sola dependencia para el beneficio de todas las dependencias.
2. **Comunitarias** - acciones adoptadas por un grupo de dependencias en torno a un tema común.
3. **Específicas de cada dependencia** - acciones para que una sola dependencia desarrolle únicamente las aptitudes de su personal.

Figura A.4: Relación entre la Estrategia y el Plan de Acción

Fuente: https://strategy.data.gov/assets/docs/draft-2019-2020-federal-data-strategy-action-plan.pdf.

Cada una de las acciones que se mencionan en el plan establece explícitamente:

- La práctica con la que se asocia la acción
- La oficina o dependencia responsable
- El plazo para su finalización
- Cómo se determinará si ha sido exitosa.

En la Tabla A.1 se incluyen varios ejemplos de muestra. El gobierno de los EE. UU. prevé que se publicará una versión final del plan de acción en septiembre de 2019.

Tabla A.1: Ejemplos de acciones

Acción	Descripción	Fecha límite
Mejorar los recursos de datos para la investigación y el desarrollo de la IA	Mejorar la documentación de los inventarios de datos y modelos para permitir su detección y utilización, y dar prioridad a las mejoras en el acceso y la calidad de los datos y modelos de IA, basándose en la retroalimentación de los usuarios de la comunidad de investigadores de IA.	Febrero de 2020
Desarrollar un marco ético de datos	Establecer un marco coherente para evaluar las repercusiones y concesiones éticas asociadas a la gestión y el uso de los datos.	Noviembre de 2019
Identificar oportunidades para aumentar las competencias del personal en materia de datos	Identificar las competencias cruciales en materia de datos para cada dependencia Evaluar la capacidad actual del personal en cuanto a competencias cruciales en materia de datos. Elaborar un plan inicial para atender las lagunas entre las necesidades de competencias cruciales en materia de datos y la capacidad actual.	Mayo de 2020

Fuente: https://strategy.data.gov/action-plan.

Programa de Políticas Públicas del Instituto Alan Turing (Reino Unido)

Problema

Los gobiernos tienen acceso a grandes cantidades de datos. Esos datos pueden ser recopilados mediante procesos de estadísticas oficiales, por ejemplo, a través de encuestas. Sin embargo, también se genera una gran cantidad de datos a través de las interacciones y transacciones diarias del gobierno con los ciudadanos y las empresas. La IA tiene un enorme potencial para que los gobiernos aprovechen estas vastas cantidades de datos y proporcionen a los funcionarios conocimientos sin precedentes, que les permitan mejorar los procesos de elaboración de políticas y hacer más eficientes los servicios públicos (Margetts y Dorobantu, 2019). Por ejemplo, la IA podría:

- proporcionar pronósticos y predicciones más precisas, lo cual permitiría a los gobiernos planificar más eficazmente y dirigir los recursos y servicios donde más se necesitan
- adaptar los servicios públicos a las necesidades de los usuarios, lo cual permitiría a los gobiernos adaptar los servicios que ofrecen a las circunstancias individuales
- simular sistemas complejos, desde operaciones militares hasta mercados inmobiliarios, lo cual brindaría a los gobiernos la oportunidad de experimentar con opciones de políticas y detectar consecuencias no deseadas antes de comprometerse con nuevas medidas.

A pesar de la promesa que guarda la IA para la elaboración de políticas y la prestación de servicios, el desarrollo y la retención de conocimientos especializados internos sobre las tecnologías y los enfoques emergentes y sus aplicaciones en el sector público, es un desafío para todos los gobiernos. En el pasado, los gobiernos se han esforzado por adoptar soluciones mucho más sencillas, como los sistemas de nómina electrónica o los sistemas de reserva de citas en línea (Margetts y Dorobantu, 2019).

Además, recurrir a expertos externos conlleva su propio conjunto de dificultades, como la gestión eficaz de los contratos, el traslado de datos a través de los límites de las dependencias u organizaciones, y cuestiones culturales más amplias en torno a las diferentes prioridades y formas de trabajo de las mismas. Cuando compran algoritmos y tecnologías basadas en IA a proveedores externos, los gobiernos batallan para determinar la relación calidad-precio y evaluar el desempeño de estos complejos productos.

Respuesta

El Instituto Alan Turing es el instituto nacional del Reino Unido para el estudio de la ciencia de datos y la inteligencia artificial. Fundado como una organización benéfica en 2015, con un enfoque en la ciencia de datos, el Instituto añadió la IA a su cometido en 2017.[26] En el recuadro A.8 se exponen los objetivos del Instituto Alan Turing.

[26] www.turing.ac.uk/about-us.

Recuadro A.8: Los objetivos del Instituto Alan Turing

- **Avanzar en la investigación de clase mundial y aplicarla a los problemas del mundo real**: innovar y desarrollar investigaciones de clase mundial en ciencia de datos e inteligencia artificial que respalden los desarrollos teóricos de próxima generación y se apliquen a los problemas del mundo real, fomentando la creación de nuevos negocios, servicios y empleos.
- **Entrenar a los líderes del futuro:** proporcionar capacitación a las nuevas generaciones de dirigentes de la ciencia de datos y la IA con la amplitud y profundidad necesarias de los conocimientos técnicos y éticos que satisfagan las crecientes necesidades industriales y sociales del Reino Unido.
- **Dirigir la conversación pública:** a través de investigaciones encaminadas a fijar la agenda, la participación pública y la asesoría técnica de expertos, impulsar ideas nuevas e innovadoras que tengan una influencia significativa en la industria, el gobierno, la reglamentación o las opiniones de la sociedad, o que repercutan en la forma en que se llevan a cabo las investigaciones sobre ciencia de datos e inteligencia artificial.

Fuente: www.turing.ac.uk/about-us.

El Instituto Alan Turing colabora con 13 universidades de investigación líderes de todo el Reino Unido, así como con el Consejo de Investigación de Ingeniería y Ciencias Físicas [*Engineering and Physical Sciences Research Council*]. Cerca de 500 académicos ocupan puestos en el Instituto, lo que le da un acceso sin precedentes a conocimientos especializados en diversas disciplinas, desde informática y matemáticas hasta ciencias sociales y filosofía.[27]

La misión del Instituto Alan Turing es "lograr grandes avances en la ciencia de datos y en la investigación de la inteligencia artificial para tener un mundo mejor". Como paso importante hacia el cumplimiento de su misión, el Instituto puso en marcha un programa de investigación sobre políticas públicas en mayo de 2018. El programa trabaja junto con los encargados de elaborar las políticas para desarrollar investigaciones, herramientas y técnicas de IA que tengan un impacto positivo en la vida del mayor número posible de personas.[28] Los retos del Programa de Políticas Públicas se exponen en el Recuadro A.9.

Recuadro A.9: Los desafíos del Programa de Políticas Públicas del Instituto Alan Turing

- **Utilizar la ciencia de datos y la Inteligencia Artificial para informar la elaboración de políticas.** En un mundo de medidas políticas cambiantes e interrelacionadas, la ciencia de datos y la IA pueden proporcionar a los encargados de formular políticas, conocimientos sin precedentes que van desde la identificación de las prioridades de las políticas mediante el modelado de sistemas y situaciones hipotéticas complejas, hasta la evaluación de resultados políticos difíciles de medir. El objetivo

[27] www.turing.ac.uk/about-us.
[28] www.turing.ac.uk/research/research-programmes/public-policy.

Recuadro A.9: Los desafíos del Programa de Políticas Públicas del Instituto Alan Turing *(Cont.)*

del Programa de Políticas Públicas es dotar a los encargados de formular políticas de todos los niveles de gobierno, de las herramientas que necesitan no sólo para diseñar políticas públicas eficaces, sino también para rastrear y medir los impactos de las mismas.

- **Mejorar la prestación de los servicios públicos.** Actualmente, los gobiernos son los principales poseedores de datos que la ciencia de datos y la IA pueden aprovechar para mejorar el diseño y la prestación de los servicios públicos. El Programa de Políticas Públicas reúne a investigadores y responsables de formular políticas con el fin de desarrollar formas innovadoras de prestar servicios públicos. El objetivo del programa es cambiar la vida cotidiana para mejor, desde la asignación de recursos de la manera más justa y transparente, hasta el diseño de servicios públicos personalizados, adaptados a las necesidades y situaciones individuales de las personas.

- **Construir fundamentos éticos para el uso de la ciencia de datos y la IA en la elaboración de políticas.** Comprender las repercusiones éticas y sociales de la ciencia de datos y la IA es parte fundamental del desarrollo de las tecnologías y las estrategias de Inteligencia Artificial. El Programa de Políticas Públicas trabaja con los responsables de formular políticas para desarrollar los fundamentos éticos del uso de la ciencia de datos y la IA en el sector público, con el fin de asegurar los beneficios y abordar los riesgos que plantean.

- **Contribuir a las políticas que rigen el uso de la ciencia de datos y la IA.** Los efectos de la ciencia de datos y la IA en la sociedad ya se están sintiendo, y su impacto sólo aumentará en los próximos años. El Programa de Políticas Públicas trabaja en conjunto con los gobiernos y los organismos reguladores para crear leyes bien elaboradas y una reglamentación sensata, con el fin de garantizar que el impacto de las tecnologías que surgen de estos poderosos campos sea lo más beneficioso y equitativo posible.

Fuente: www.turing.ac.uk/research/research-programmes/public-policy.

Un equipo central de investigadores con formación en ciencias sociales supervisa las actividades del Programa de Políticas Públicas. Su experiencia académica va desde la filosofía y el derecho, hasta la economía y las relaciones internacionales.

El programa se encuentra en una posición única para trabajar junto con los responsables de formular políticas en el uso de la ciencia de datos y la IA para el bien común. En el Reino Unido, el programa ofrece un punto de contacto único para que el gobierno recurra a los principales expertos académicos del país en ciencias de datos e IA. Al ofrecer asesoría imparcial, los investigadores académicos independientes ocupan un lugar privilegiado para ayudar a los gobiernos a maximizar el potencial de la ciencia de datos y la IA para resolver los problemas de políticas públicas (Margetts y Dorobantu, 2019).

A través del Programa de Políticas Públicas, los funcionarios públicos tienen acceso a la creciente red de académicos del Instituto Alan Turing. El equipo central de investigadores del programa pone en contacto a los responsables de formular políticas con los académicos para examinar los problemas a los que se enfrentan, y determinar las mejores metodologías e instrumentos disponibles para resolverlos. Los investigadores académicos pueden entonces trabajar en el Instituto para elaborar los instrumentos pertinentes o pueden integrarse en equipos del sector público para llevar a cabo un proyecto. Al final de cada colaboración, el programa entrega las herramientas y técnicas desarrolladas a los funcionarios públicos, que a su vez son capacitados por los académicos para asumir la propiedad de los proyectos.

Resultados e impacto

El Instituto Alan Turing es una parte integral de la perspectiva sobre la IA que el gobierno del Reino Unido ha adoptado, la cual se estableció en el Acuerdo del Sector de IA como parte de la Estrategia Industrial del país (Gov.UK [*Gobierno del Reino Unido*], 2019b). Los organismos del sector público pueden consultar el Programa de Políticas Públicas de Turing para obtener asesoría fiable e independiente sobre la IA y la ciencia de datos, incluidas las cuestiones éticas. La Comisión Europea ha señalado el valor de esta clase de institutos y programas en su informe normativo emblemático sobre IA, denominado *Inteligencia Artificial*: *Una Perspectiva Europea*.[29]

A pesar de que el Programa de Políticas Públicas no existe sino desde hace poco más de un año, el impacto que ha tenido hasta el momento ha sido considerable. El programa ha brindado asesoría y orientación a cientos de encargados de formular políticas, en representación de más de 80 organizaciones, desde los gobiernos locales y las fuerzas policiales hasta los departamentos del gobierno central, los organismos reguladores y las organizaciones internacionales. El programa también ha contribuido a iniciativas de políticas fundamentales en el Reino Unido, entre ellas, el establecimiento del Centro de Ética e Innovación de Datos del Gobierno del Reino Unido[30] y la Estrategia de Innovación Tecnológica del Gobierno.

Además de su función como asesor, el Programa de Políticas Públicas alberga más de 20 proyectos de investigación plurianuales, en los que participan más de 60 investigadores académicos de 10 universidades. Cada proyecto aborda un desafío de política específico y está dirigido por un académico de alto nivel. Los proyectos de investigación abarcan una amplia gama de temas, desde la medición y la lucha contra los discursos de incitación al odio en línea, hasta el aumento del número de mujeres en la ciencia de datos y la IA, así como la identificación de las políticas que los países en desarrollo deben priorizar para alcanzar los objetivos de desarrollo sostenible de las Naciones Unidas.

Algunos de los proyectos del programa ya están teniendo un impacto directo en el mundo de las políticas. En 2019, el Programa de Políticas Públicas del Instituto se asoció con la Oficina de Inteligencia Artificial del Gobierno del Reino Unido y el Servicio Digital del Gobierno para generar guías sobre el diseño y la aplicación responsables de los sistemas de IA en el sector público. La guía, *Comprender los aspectos éticos y de seguridad de la Inteligencia Artificial*

[29] https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-european-perspective.
[30] www.turing.ac.uk/research/publications/dcms-consultation-centre-data-ethics-and-innovation.

[*Understanding Artificial Intelligence Ethics and Safety*],[31] fue escrita por investigadores del Instituto Turing y publicada por el Ministro de Aplicación del Reino Unido en junio de 2019. Representa la guía más completa del mundo en materia de ética y seguridad de la IA para el sector público, e identifica los posibles daños causados por los sistemas de IA, y propone medidas concretas y operativas para contrarrestarlos.

El Programa de Políticas Públicas del Instituto Turing también colabora con los organismos regulatorios del Reino Unido en la explicabilidad y la transparencia de la IA. El programa también está trabajando con la Oficina del Comisionado de Información (ICO, por sus siglas en inglés) para desarrollar una guía que ayude a las dependencias a explicar las decisiones de IA a las personas que se vean afectadas por éstas. En junio se publicó un informe provisional y la guía final se elaborará en el tercer trimestre de 2019 (ICO, 2019). El programa también anunció recientemente una nueva colaboración con la Autoridad de Conducta Financiera en un proyecto de investigación que examinará los usos actuales y futuros de la IA en todo el sector de los servicios financieros, analizará las cuestiones éticas y regulatorias que se plantean en este contexto y prestará asesoramiento sobre posibles estrategias para abordarlas.[32]

Desafíos y lecciones aprendidas

El programa es un asesor y colaborador de confianza de los organismos reguladores y las oficinas gubernamentales. La capacidad de mantener su independencia en esa función es de suma importancia, por lo que el programa depende únicamente de los fondos públicos para desarrollar su labor. El equipo de investigadores del programa cuenta con el apoyo de una combinación de financiamiento básico del Instituto Alan Turing y de subvenciones de la Oficina de Investigación e Innovación del Reino Unido, un organismo nacional de financiamiento que invierte en ciencia e investigación. Si bien este modelo de financiamiento ha resultado exitoso, en ocasiones puede limitar la capacidad del programa para trabajar en proyectos que quedan fuera del alcance de sus subvenciones.

En el caso de los proyectos de investigación que quedan fuera de las condiciones de subvención, el programa pasa por procesos de contratación pública. Sin embargo, las adquisiciones del sector público no siempre se adaptan bien a los proyectos de investigación de IA. Los plazos de los proyectos son en ocasiones demasiado cortos, sobre todo cuando los departamentos gubernamentales necesitan entregar resultados concretos en plazos muy ajustados. En el ámbito académico, es preciso contratar investigadores posdoctorales para cada nuevo proyecto, un proceso largo en contraste con las consultorías privadas, que tienen empleados que pueden ser asignados inmediatamente a un nuevo proyecto. Los contratos gubernamentales también pueden ir acompañados de condiciones estrictas que socavan el avance de la carrera académica, que está estrechamente ligada a la publicación de los resultados de las investigaciones. Por ejemplo, algunos contratos de adquisición no permiten ninguna forma de publicación, lo que disminuye el atractivo del trabajo para los investigadores académicos.

Ninguno de estos desafíos es insuperable. En reconocimiento de la necesidad de un nuevo marco para el diseño, la adquisición y el despliegue efectivos y responsables de IA por parte

31 www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety.
32 www.turing.ac.uk/news/new-collaboration-fca-ethical-and-regulatory-issues-concerning-use-ai-financial-sector.

del sector público, la Oficina para la IA del Gobierno del Reino Unido se asoció este año con el Foro Económico Mundial para diseñar en conjunto las directrices para la adquisición de IA (Gerdon, 2019).

El Programa de Políticas Públicas está estudiando otras vías para financiar la ciencia de datos y la investigación en materia de IA, así como para estimular a las instituciones de investigación a que desempeñen un papel más importante en la innovación del sector público. Entre las soluciones posibles figuran la creación de un fondo gubernamental para la innovación o el desarrollo de nuevos modelos de financiamiento que permitan a los departamentos y dependencias gubernamentales, colaborar con socios académicos.

Anexo B. Glosario y Códigos de países

Glosario

Aprendizaje Automático (AA): Un subconjunto de la IA y un método en el que las máquinas aprenden a hacer predicciones en nuevas situaciones basadas en datos históricos. El Aprendizaje Automático consiste en un conjunto de técnicas que permiten que las máquinas aprendan de manera automatizada, sin instrucciones explícitas de un humano, basándose en patrones e inferencias. Los métodos de Aprendizaje Automático enseñan con frecuencia a las máquinas a alcanzar un resultado, proporcionándoles muchos ejemplos de resultados correctos denominado "entrenamiento". En otro método, los humanos definen un conjunto de reglas generales y normalmente dejan que la máquina aprenda por sí misma a través del proceso de ensayo y error.

Aprendizaje no supervisado: Es una forma de AA que implica proporcionar al sistema datos que no contienen las variables de salida. El aprendizaje no supervisado se utiliza principalmente para identificar patrones en los datos.

Aprendizaje por refuerzo: Es una forma de AA que funciona al hacer que un agente (computadora) complete una tarea por medio de la interacción con un entorno. Con base en dichas interacciones, el entorno proporcionará una retroalimentación que hará que el agente adapte su comportamiento. En otros términos, el agente *aprende* a través del *ensayo y error;* el entorno penaliza el error y recompensa el éxito. Luego, conforme pasa el tiempo ajusta automáticamente su comportamiento, lo cual da como resultado acciones más refinadas.

Aprendizaje Profundo: Es una forma de Aprendizaje Automático cuyo diseño está inspirado en la biología del cerebro humano. El Aprendizaje Profundo funciona al exponer redes neuronales artificiales de múltiples capas (véase la definición en apartados anteriores) a grandes cantidades de datos.

Aprendizaje supervisado: Es una forma de AA que implica proporcionar al sistema datos que contienen tanto las variables de entrada (*características*) como las variables de salida (*etiqueta, respuesta*) para entrenar al sistema.

Aumento de la Inteligencia Artificial (aumento de la inteligencia): Se refiere a los casos en que los humanos y las máquinas aprenden unos de otros y redefinen la amplitud y la profundidad de lo que hacen juntos.

Automatización Robótica de Procesos (RPA, por sus siglas en inglés): Una tecnología de automatización de procesos comerciales que automatiza tareas manuales que están

basadas en reglas, son estructuradas y repetitivas utilizando softwares robóticos, también denominados bots. Las herramientas de RPA esquematizan un proceso para que un robot lo siga, lo que le permite operar en lugar de un humano.

Característica de datos: Una variable explicativa en un conjunto de datos que se utiliza para hacer predicciones

Efecto Inteligencia Artificial (efecto IA): El fenómeno mediante el cual "los investigadores de la Inteligencia Artificial (IA) logran un hito que durante mucho tiempo se pensó que significaba el logro de la verdadera inteligencia artificial, por ejemplo, ganarle a un humano en el ajedrez, y de repente se degrada a IA no *verdadera*."

Entrenamiento del aprendizaje automático: Fase en la que se expone a un sistema de IA a los datos de los que aprende aplicando modelos estadísticos.

Etiquetado de datos: Implica identificar los elementos que un sistema de IA tratará de predecir. También se puede denominar *respuesta*, *resultado* o *salida*.

Facetas de la innovación: La OPSI ha identificado cuatro facetas de la innovación del sector público. Éstas son:

- La *innovación orientada a la misión* establece un resultado claro y un objetivo general para lograr una misión específica.
- La *innovación orientada al mejoramiento* actualiza las prácticas, logra eficiencias y mejores resultados y se basa en las estructuras existentes.
- La *innovación adaptable* analiza y pone a prueba nuevos enfoques a fin de responder a un entorno operativo cambiante.
- La *innovación anticipada* explora y se involucra en problemas emergentes que podrían determinar las prioridades y compromisos futuros.

Generalización: Se refiere a la capacidad de un modelo de IA para hacer predicciones correctas sobre nuevos datos nunca antes vistos, en contraposición a los datos utilizados para entrenar el modelo.

Generalización del aprendizaje automático: Una fase posterior a la validación en la que se despliega un sistema de aprendizaje automático en un entorno real.

Ingeniería de características: El proceso de determinar qué características podrían ser útiles para el entrenamiento de un modelo.

Inteligencia Artificial (IA): No existe una definición universalmente aceptada para la IA. IA es un término genérico acuñado en la década de 1950 para hablar de sistemas, máquinas o computadoras que parecen imitar la forma de pensar de los seres humanos.

Actualmente, éste puede usarse para describir diferentes cosas:

- La IA como campo de investigación que abarca el diseño de sistemas inteligentes y el desarrollo de métodos y técnicas conexos, pero también incluye muchas otras cuestiones, tales como el impacto ético y social de dichos sistemas.
- La IA como tecnología (es decir, la aplicación de este acervo de conocimientos para resolver problemas de la vida real).

Inteligencia Artificial Estrecha (IAE): Es un término utilizado para reconocer que todas las aplicaciones actuales de la IA pueden utilizarse eficazmente para realizar algunas tareas específicas, pero que no suelen ser buenas cuando se utilizan en otro contexto.

Inteligencia artificial general (AGI, por sus siglas en inglés): Se refiere a la idea de que la inteligencia humana general, que abarca diferentes dominios y capacidades, podría ser igualada o incluso superada por las máquinas. Algunas investigaciones utilizan el término "Súper Inteligencia Artificial", o "superinteligencia", para referirse a sistemas hipotéticos de IA general que podrían superar por mucho las capacidades de los humanos

Interfaz de programación de aplicaciones (API, por sus siglas en inglés): Consiste en un conjunto de definiciones, protocolos y herramientas para crear software de aplicación. En pocas palabras, se trata de un conjunto de métodos que permiten que los componentes de software interactúen entre sí, posibilitando, por ejemplo, que los usuarios copien y peguen texto u otros tipos de datos de una aplicación a otra. Existen muchos tipos diferentes de API para sistemas operativos, aplicaciones o sitios web. Una buena API facilita el desarrollo de un programa, ya que proporciona todos los bloques de información, los cuales posteriormente son ensamblados por un programador

Neurona (o nodo) artificial: Un modelo matemático inspirado en la biología de los cerebros humanos. Suele tener muchas entradas y una sola salida unidas por una función matemática, también llamada *función de activación* o *función de transferencia*. Cuando se cuenta con varias neuronas enlazadas entre sí, se puede crear una *red neuronal artificial*.

Normalización de datos: la conversión de los valores de su estado *bruto* a un rango de valores estándar.

Planificación automatizada: Se refiere a la capacidad que tienen las máquinas para diseñar de manera automática y autónoma medidas o estrategias para lograr un objetivo, incluyendo la anticipación de los efectos de diferentes enfoques.

Precisión: Es la forma más común en que se mide el desempeño del Aprendizaje Automático. Es una expresión de la frecuencia con la que el sistema de IA proporciona una respuesta correcta, en contraposición a un falso positivo o un falso negativo.

Procesamiento del Lenguaje Natural (PLN): Se refiere a la capacidad que tienen las computadoras para manipular e interpretar el lenguaje humano y realizar diversas tareas como la traducción o el análisis de textos. Además de procesar el lenguaje existente, las aplicaciones de PLN también pueden generar un nuevo lenguaje hablado o escrito.

Reconocimiento de sonidos vocales: Se refiere a la capacidad que tienen las computadoras para analizar archivos de audio con el fin de reconocer e interpretar el lenguaje hablado.

Redes Neuronales Artificiales (RNA): Un conjunto de neuronas artificiales enlazadas, diseñadas para realizar cálculos para completar tareas específicas. La forma más simple de RNA está compuesta por tres capas de neuronas: una *capa de entrada*, una *capa oculta* y una *capa de salida*. Todas las neuronas de la capa de entrada están enlazadas con todas las neuronas de la capa oculta, que a su vez están enlazadas con todas las neuronas de la capa de salida.

Sesgo:

- El *sesgo de muestreo* es un sesgo en el proceso de recopilación de datos (p. ej., el sobremuestreo o submuestreo de grupos específicos).
- El *sesgo estadístico* se refiere sobre todo a un modelo que genera sistemáticamente un error de predicción cuando se compara con el resultado esperado

Sistema de Inteligencia Artificial: Un sistema basado en máquinas que puede, para un conjunto determinado de objetivos definidos por el ser humano, hacer predicciones, recomendaciones o tomar decisiones que influyen en entornos reales o virtuales. Los sistemas de IA están diseñados para funcionar con diversos niveles de autonomía. Además, las IA son "máquinas que realizan funciones cognitivas similares a las de los seres humanos".

Sistemas basados en el conocimiento: Registran y almacenan hechos en una "base de conocimientos", y posteriormente utilizan un "motor de inferencias" para inferir información de la base de conocimiento con el fin de resolver problemas, con frecuencia a través de reglas programadas del tipo SI…ENTONCES.[1] Por ejemplo, los "sistemas expertos" codifican los conocimientos de los expertos para ayudar a resolver problemas a través de reglas del tipo SI…ENTONCES.

Sobreajuste: Se refiere a los casos en que un algoritmo es demasiado específico, al grado que registra y se centra demasiado en el ruido y las anomalías.

Subajuste: Se refiere a las situaciones en que un sistema de Aprendizaje Automático no puede registrar la información subyacente contenida en los datos.

Validación del aprendizaje automático: Es la fase posterior al entrenamiento en la que los operadores se aseguran de que el sistema esté correctamente entrenado para resolver el problema inicialmente definido, cotejando los resultados del sistema con los datos de muestra.

Visión artificial: Se refiere a la capacidad de la IA de procesar y sintetizar datos visuales (por ejemplo, detectar y clasificar objetos con base en imágenes o videos) y realizar tareas como el reconocimiento facial y la interpretación de escenas. Además de procesar los datos visuales existentes, algunas aplicaciones implican la construcción de datos visuales, como la creación de modelos en 3D a partir de imágenes en 2D.

Nota: Varias de estas definiciones proceden de fuentes externas. Estas fuentes se citan en el cuerpo principal del informe.

[1] https://searchcio.techtarget.com/definition/knowledge-based-systems-KBS.

Códigos de países

ARG	Argentina
AUS	Australia
AUT	Austria
BEL	Bélgica
BGR	Bulgaria
BRA	Brasil
CAN	Canadá
CHE	Suiza
CHL	Chile
COL	Colombia
CRI	Costa Rica
CZE	República Checa
DEU	Alemania
DNK	Dinamarca
DOM	República Dominicana
ESP	España
EST	Estonia
FIN	Finlandia
FRA	Francia
GBR	Reino Unido
GRC	Grecia
GTM	Guatemala
HND	Honduras
HUN	Hungría
IDN	Indonesia
IRL	Irlanda
ISL	Islandia
ISR	Israel
ITA	Italia
JPN	Japón
KEN	Kenia
KOR	Corea
KWT	Kuwait
LBN	Líbano
LTU	Lituania
LUX	Luxemburgo
LVA	Letonia
MAR	Marruecos
MEX	México
MLT	Malta
NOR	Noruega
NLD	Países Bajos
NZL	Nueva Zelanda
PAN	Panamá
PER	Perú
PHL	Filipinas
POL	Polonia
PRT	Portugal
QAT	Qatar
ROU	Rumania
SAU	Arabia Saudita
SGP	Singapur
SLV	El Salvador
SVK	República Eslovaca
SVN	Eslovenia
SWE	Suecia
TUN	Túnez
TUR	Turquía
URY	Uruguay
USA	Estados Unidos
ZAF	Sudáfrica
ZMB	Zambia

Referencias

Agrawal, A., J. Gans y A. Goldfarb (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Review Pres, Boston, MA.

AI NOW (2018), "After a Year of Tech Scandals, Our 10 Recommendations for AI", *Medium,* 6 de diciembre, https://medium.com/@AINowInstitute/after-a-year-of-tech-scandals-our-10-recommendations-for-ai-95b3b2c5e5.

Alom, Z. Md, T.M. Taha, C Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, B.C. Van Essen, A.A.S. Awaal y V.K. Asari (2018), "The history began from AlexNet: A comprehensive survey on deep learning approaches", https://arxiv.org/ftp/arxiv/papers/1803/1803.01164.pdf.

Anastasopoulos, L.J. y A.B. Whitford (2019), "Machine learning for public administration research, with application to organizational reputation", *Journal of Public Administration Research and Theory*, Vol. 29/3, pp. 491-510, https://doi.org/10.1093/jopart/muy060.

Andrews, M. (2018), *How Do Governments Build Capabilities to Do Great Things? The Oxford Handbook of the Politics of Development*, Oxford University Press, https://doi.org/10.1093/oxfordhb/9780199845156.013.34.

AuroraAI (2019), *AuroraAI – Towards a Humancentric Society*, https://vm.fi/documents/10623/1464506/AuroraAI+development+and+implementation+plan+2019%E2%80%932023.pdf.

Balaram, B., T. Greenham y J. Leonard (2018), *Artificial Intelligence: Real Public Engagement*, Londres, RSA, www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence---real-public-engagement.pdf.

Balestra, C. y L. Fleischer (2018), "Diversity statistics in the OECD: How do OECD countries collect data on ethnic, racial and indigenous identity?", *OECD Statistics Working Papers* 2018/09, OECD Publishing, París, www.oecd-ilibrary.org/docserver/89bae654-en.pdf?expires=1571087561&id=id&accname=guest&checksum=8328FF28C0783A6837863C112647E63F.

Bansak, K., J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence y J. Weinstein (2018), "Improving refugee integration through data-driven algorithmic assignment", *Science*, Vol. 359/6373, pp. 325-329, http://science.sciencemag.org/content/359/6373/325.

BBC News (2019), "Could an algorithm help prevent murders?", 24 de junio, www.bbc.com/news/stories-48718948.

Bostrom, N. (2014), *Superintelligence*, Oxford University Press.

Bryman, A. (2016), *Social Research Methods*, Oxford University Press.

Carrasco, M., S. Mills, A. Whybrew y A. Jura (2019), "The citizen's perspective on the use of AI in government", *BCG*, 1 de marzo, www.bcg.com/publications/2019/citizen-perspective-use-artificial-intelligence-government-digital-benchmarking.aspx.

Carter, S. y M. Nielsen (2017), "Using Artificial Intelligence to augment human intelligence", *Distill*, 4 de diciembre, https://distill.pub/2017/aia.

Case, N. (2018), "How to become a centaur", *JoDs*, 8 de enero, https://jods.mitpress.mit.edu/pub/issue3-case.

CIPL (2019), "Regulatory sandboxes in data protection: Constructive engagement and innovative regulation in practice", Centre for Information Policy Leadership, Washington, DC, www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_white_paper_on_regulatory_sandboxes_in_data_protection_-_constructive_engagement_and_innovative_regulation_in_practice__8_march_2019_.pdf.

CSSF (2018), *Artificial Intelligence: Opportunities, Risks and Recommendations for the Financial Sector*, Luxembourgo, Commission de Surveillance du Secteur Financier, www.abbl.lu/content/uploads/2018/12/CSSF_White_Paper_Artificial_Intelligence_20 1218.pdf.

Dastin, J. (2018), "Amazon scraps secret AI recruiting tool that showed bias against women", *Reuters*, 10 de octubre, www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

Dencik, L., A. Hintz, J. Redden y H. Warner (2018), *Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services. Project Report*, Data Justice Lab/Cardiff University/Open Society Foundations, https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf.

Desouza, K.C. (2018), *Delivering Artificial Intelligence in Government: Challenges and Opportunities*, IBM Center for the Business of Government, Washington, DC, www.businessofgovernment.org/sites/default/files/Delivering%20Artificial%20Intelligence%20in%20Government.pdf.

Dickson, B. (2019), "What happens when you combine neural networks and rule-based AI?", *TechTalks*, 5 de junio, https://bdtechtalks.com/2019/06/05/mit-ibm-hybrid-ai.

du Preez, D. (2018), "Professor Dame Wendy Hall – 'We need to put diversity at the centre of the AI ethics debate'", *Diginomica*, 3 de diciembre, https://diginomica.com/professor-dame-wendy-hall-we-need-to-put-diversity-at-the-centre-of-the-ai-ethics-debate.

Eggers, W.D., M. Turley y P. Kishnani (2018), "The future of regulation: Principles for regulating emerging technologies", *Deloitte Insights*, 19 de junio, www2.deloitte.com/us/en/insights/industry/public-sector/future-of-regulation/regulating-emerging-technology.html.

Eggers, W.D., Schatsky, D. y Viechnicki, P. (2017), *AI-Augmented Government: Using Cognitive Technologies to Redesign Public Sector Work*, New York, Deloitte University Press, www2.deloitte.com/content/dam/insights/us/articles/3832_AI-augmented-government/DUP_AI-augmented-government.pdf.

EU2017.ee (2017), *Declaración de Tallin sobre Gobierno Electrónico en la Reunión Ministerial durante la Presidencia de Estonia del Consejo de la UE el 6 de octubre de 2017*, https://ec.europa.eu/newsroom/document.cfm?doc_id=47559.

Comisión Europea (2019a), "Ethics guidelines for trustworthy AI", Comisión Europea, Bruselas, https://ec.europa.eu/knowledge4policy/publication/ethics-guidelines-trustworthy-ai_en.

Comisión Europea (2019b), *A Definition of AI: Main Capabilities and Disciplines*, definición desarrollada para los fines de los entregables de IA HLEG, Grupo Independiente sobre Inteligencia Artificial establecido por la Comisión Europea, Bruselas, https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines.

Comisión Europea (2018a), *Artificial Intelligence: A European Perspective*, Comisión Europea, Bruselas, https://ec.europa.eu/jrc/en/publication/eur-scientificand-technical-research-reports/artificial-intelligence-european-perspective.

Comisión Europea (2018b), *Artificial Intelligence for Europe*, Comisión Europea, Bruselas, https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-237-F1-EN-MAIN-PART-1.PDF.

Forbes Insights (2011), *Global Diversity and Inclusion: Fostering Innovation Through a Diverse Workforce*, Forbes, Nueva York, https://i.forbesimg.com/forbesinsights/StudyPDFs/Innovation_Through_Diversity.pdf.

Frank, M.R., D. Wang, M. Cebrian y I. Rahwan (2019), "The evolution of citation graphs in artificial intelligence research", *Nature Machine Intelligence*, Vol. 1, pp. 7985, www.nature.com/articles/s42256-019-0024-5.

Gardner, J. (1983, 2011), *Frames of Mind: The Theory of Multiple Intelligences*, Basic Books, Nueva York.

Gerdon, S. (2019), "How the public sector can procure AI-powered solutions more effectively and responsibly", *DCMS blog*, 24 de julio, https://dcmsblog.uk/2019/07/how-the-public-sector-can-procure-ai-powered-solutions-more-effectively-and-responsibly.

Gomes de Sousa, W., E.R. Pereira De Melo, P.H. De Souza Bermejo, R.A. Sousa Farias y A. Oliveira Gomes (2019), "How and where is artificial intelligence in the public sector going? A literature review and research agenda", *Government Information Quarterly*, In Press, https://doi.org/10.1016/j.giq.2019.07.004.

Goodfellow, I, Y. Bengio y A. Courville (2016), *Deep Learning*, MIT Press, Cambridge, MA, www.deeplearningbook.org.

Gobierno de Canadá (2019a), "Ensuring responsible use of artificial intelligence to improve government services for Canadians", Comunicado de prensa, 4 de marzo, www.canada.ca/en/treasury-board-secretariat/news/2019/03/ensuring-responsible-use-of-artificial-intelligence-to-improve-government-services-for-canadians.html.

Gobierno de Canadá (2019b), "Government of Canada creates Advisory Council on Artificial Intelligence", Comunicado de prensa, 14 de mayo, www.canada.ca/en/innovation-science-economic-development/news/2019/05/government-of-canada-creates-advisory-council-on-artificial-intelligence.html.

Gobierno del Reino Unido (2019a), "Leading experts appointed to AI Council to supercharge the UK's artificial intelligence sector", Comunicado de prensa, 16 de mayo, www.gov.uk/government/news/leading-experts-appointed-to-ai-council-tosupercharge-the-uks-artificial-intelligence-sector.

Gobierno del Reino Unido (2019b), *AI Sector Deal*, Documentos sobre políticas, 21 de mayo, www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal.

Grabianowski, E. (2006), "How speech recognition works", *HowStuffWorks.com*, 10 de noviembre, https://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition2.htm.

Grace, K., J. Salvatier, A. Dafoe, B. Zhang y O. Evans (2018), "Viewpoint: When will AI exceed human performance? Evidence from AI experts", *Journal of Artificial Intelligence Research*, Vol. 62/2018, pp. 729-754, https://arxiv.org/abs/1705.08807.

Graham, T. (2019), "Barcelona is leading the fightback against smart city surveillance", *Wired*, 18 de mayo, www.wired.co.uk/article/barcelona-decidim-ada-colau-francesca-briadecode.

Greenwood, M. (2019), "Canada's new Federal Directive makes ethical AI a national issue", *Techvibes*, 8 de marzo, https://techvibes.com/2019/03/08/canadas-new-federal-directive-makes-ethical-ai-a-national-issue.

Haugeland, J. (1985), *Artificial Intelligence: The Very Idea*, https://philpapers.org/rec/HAUAIT.

Herd, P. y D.P. Moynihan (2018), *Administrative Burden: Policymaking by Other Means*. Russell Sage Foundation, Nueva York, www.jstor.org/stable/10.7758/9781610448789.

IBM Center for the Business of Government (2019), *More Than Meets the AI*, New York/Washington, DC, Partnership for Public Service/IBM Center for the Business of Government, www.businessofgovernment.org/sites/default/files/More%20Than%20Meets%20AI%20Part%20II_0.pdf.

ICO (2019), *Project ExplAIn: Interim Report*, Information Commissioner's Office, Wilmslow, Reino Unido, https://ico.org.uk/media/2615039/project-explain-20190603.pdf.

IDIA (2019), *Artificial Development in International Development: A Discussion Paper*, International Development Innovation Alliance/AI & Development Working Group, https://static1.squarespace.com/static/5b156e3bf2e6b10bb0788609/t/5d3f283a3ee5d60001fcf184/1564420165214/AI+and+international+Development_FNL.pdf.

Kattel, R. (2019), "Do coders need a code of conduct?", *New Statesmen*, 6 de junio, www.newstatesman.com/spotlight/emerging-technologies/2019/06/do-coders-need-code-conduct.

Kortz, M. y F. Doshi-Velez (2017), *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center, Cambridge, MA, https://cyber.harvard.edu/publications/2017/11/AIExplanation.

Janssen, M. y J. van den Hoven (2015), "Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?", *Government Information Quarterly*, Vol. 32/4, pp. 363-368, https://doi.org/10.1016/j.giq.2015.11.007.

Laster, Y. (2018), *Developing a Strategy for Digital Transformation: Challenges & Insights*, presentación en la reunión del Grupo de Trabajo de la OCDE sobre Integración de Políticas Medioambientales y Económicas, marzo.

Lember, V., T. Brandsen y P. Tõnurist (2019), "The potential impacts of digital technologies on co-production and co-creation", *Public Management Review*, Vol. 21/11, pp. 1665-1686, www.tandfonline.com/doi/full/10.1080/14719037.2019.1619807.

Leslie, D. (2019), "Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector", Instituto Alan Turing, Londres, https://doi.org/10.5281/zenodo.3240529.

Lewin, A. (2019), "Shiny moonshot technology will not save healthcare — yet". *Sifted*, 10 June, https://sifted.eu/articles/health-tech-startups-europe-doctolib-kry-accurx.

Lighthill, J. (1973), *Artificial Intelligence: A General Survey*, www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm.

Mao, J., C. Gan, P. Kohli, J.B. Tenenbaum y J. Wu (2018), "The neuro-symbolic concept learner: Interpreting scenes, words, and sentences From natural supervision, documento publicado en ICLR 2019, https://openreview.net/pdf?id=rJgMlhRctm.

Manning, C.D., P. Raghavan y H. Schütze (2008), *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, https://nlp.stanford.edu/IR-book/information-retrieval-book.html.

Marcus, G. (2018), "In defense of skepticism about deep learning". *Medium*, 14 de enero, https://medium.com/@GaryMarcus/in-defense-of-skepticism-about-deep-learning-6e8bfd5ae0f1.

Margetts, H. y C. Dorobantu (2019), "Rethink government with AI", *Nature*, 9 de abril, www.nature.com/articles/d41586-019-01099-5.

Marr, B. (2018), "How much data do we create every day? The mind-blowing stats everyone should read", *Forbes*, 21 de mayo, www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#729e7ad460ba.

Mateos-Garcia, J. (2018), "The complex economics of artificial intelligence". *Nesta* (blog), 13 de diciembre, www.nesta.org.uk/blog/complex-economics-artificial-intelligence.

Mateos-Garcia, J. (2017), "Algorithmic fallibility and economic organisation". *Nesta* (blog), 10 de mayo, www.nesta.org.uk/blog/to-err-is-algorithm-algorithmic-fallibility-and-economic-organisation.

Mazzucato, M. (2011), *The Entrepreneurial State: Debunking Public Vs. Private Sector Myths*. Anthem Press, Londres.

McKinsey & Company (2017), *Digitally-enabled Automation and Artificial Intelligence: Shaping the Future of Work in Europe's Digital Front-Runners*, McKinsey & Company, www.mckinsey.com/~/media/mckinsey/featured%20insights/europe/shaping%20the%20future%20of%20work%20in%20europes%20nine%20digital%20front%20runner%20countries/shaping-the-future-of-work-in-europes-digital-front-runners.ashx.

Meyer, C. (2019), "How self-driving cars could make or break a green future of transportation". *Forbes*, 12 de junio, www.forbes.com/sites/christophmeyereurope/2019/06/12/a-green-future-oftransportation-how-self-driving-cars-will-be-make-or-break/#3c8961522337.

MGI (2018), *Notes from the AI Frontier: Applying AI for Social Good*, McKinsey Global Institute, www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Applying%20artificial%20intelligence%20for%20social%20good/MGI-ApplyingAI-for-social-good-Discussion-paper-Dec-2018.ashx.

Miaihle, N. y C. Hodes (2017), "Making the AI revolution work for everyone", The Future Society at the Harvard Kennedy School of Government, Cambridge, MA, http://ai-initiative.org/wp-content/uploads/2017/08/Making-the-AI-Revolution-work-for-everyone.-Report-to-OECD.-MARCH-2017.pdf.

Mikhaylov, S., M. Esteve y A. Campion (2018), "AI for the public sector: Opportunities and challenges of cross-sector collaboration", *Philosophical Transactions of the Royal Society* A, 376: 20170357.

Miller, R. (2018), *Transforming the Future (Open Access): Anticipation in the 21st Century,* Routledge, Abingdon, Reino Unido.

MMC Ventures (2019), *The State of AI 2019: Divergence*, MMC Ventures, Londres, www.mmcventures.com/wp-content/uploads/2019/02/The-State-of-AI-2019-Divergence.pdf.

Moneycontrol News (2019), "Gartner debunks five Artificial Intelligence misconceptions". *Moneycontrol*, 15 de febrero, www.moneycontrol.com/news/business/companies/gartner-debunks-five-artificial-intelligence-misconceptions-3545891.html.

Mulgan, G. (2019), "Intelligence as an outcome not an input: How can pioneers ensure AI leads to more intelligent outcomes?" *Nesta* (blog), 11 de junio, www.nesta.org.uk/blog/intelligence-outcome-not-input/?utm_source=Nesta+Weekly+Newsletter&utm_campaign=848de748ed-EMAIL_CAMPAIGN_2019_06_07_10_07&utm_medium=email&utm_term=0_d173 64114d-848de748ed-182049541.

Muller, V.C. y N. Bostrom (2014), "Future progress in Artificial Intelligence: A survey of expert opinion", in V.C. Muller (ed.), *Fundamental Issues of Artificial Intelligence*, Berlín, Biblioteca Springer/Synthese.

Nelson, R. (2019), "AI beats radiologists for accuracy in lung cancer screening", *Medscape*, 23 de mayo, www.medscape.com/viewarticle/913428.

NSTC (2016), *Preparing for the Future of Artificial Intelligence*, Oficina Ejecutiva del Presidente Comité del Consejo Nacional de Ciencia y Tecnología sobre Tecnología, Washington, DC, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

OCDE (2019a), *Artificial Intelligence in Society*, OECD Publishing, París, https://doi.org/10.1787/eedfee77-en.

OCDE (2019b), *The Future of Work: OECD Employment Handbook 2019. Highlights*, OECD Publishing, Paríis, www.oecd.org/employment/Employment-Outlook-2019Highlight-EN.pdf.

OCDE (2019c), "Using digital technologies to improve the design and enforcement of public policies", *OECD Digital Economy Papers*, No. 274, OECD Publishing, Paríis, https://dx.doi.org/10.1787/99b9ba70-en.

OCDE (2019d), *Going Digital: Shaping Policies, Improving Lives*, OECD Publishing, París, https://dx.doi.org/10.1787/9789264312012-en.

OCDE (2018a), *Science, Technology and Innovation Outlook 2018*, OECD Publishing, París, www.oecd.org/sti/oecd-science-technology-and-innovation-outlook-25186167.htm.

OCDE (2018b), *IoT Measurement and Applications*, DSTI/CDEP/CISP/MADE(2017)1/FINAL, www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/CDEP/CISP/MADE(2017)1/FINAL&docLanguage=En.

OCDE (2018c), *OECD Regulatory Policy Outlook 2018*, OECD Publishing, París. https://doi.org/10.1787/9789264303072-en.

OCDE (2017a), *Fostering Innovation in the Public Sector*, OECD Publishing, París, www.oecd.org/gov/fostering-innovation-in-the-public-sector-9789264270879-en.htm.

OCDE (2017b), *Embracing Innovation in Government: Public Trends*, OECD Publishing, París, www.oecd.org/gov/innovative-government/embracing-innovationin-government.pdf.

OCDE (2017c), *The Next Production Revolution: Implications for Governments and Business*, OECD Publishing, París, http://dx.doi.org/10.1787/9789264271036-en.

OCDE (2015a), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, París.

OCDE (2015b), *The Innovation Imperative in the Public Sector: Setting an Agenda for Action,* OECD Publishing, París, http://dx.doi.org/10.1787/9789264236561-en.

Owen, H. y K. Stathoulopoulos (2019), "How gender diverse is the workforce of AI?", Nesta, 17 de julio, www.nesta.org.uk/blog/how-gender-diverse-workforce-ai-research.

Partnership for Public Service/IBM Center for the Business of Government (2019), *More than Meets AI: Assessing the Impact of Artificial Intelligence on the Work of Government*, Washington, DC, www.businessofgovernment.org/sites/default/files/More%20Than%20Meets%20AI.pdf.

Partnership for Public Service/IBM Center for the Business of Government (2018), *The Future Has Begun*, Washington, DC, www.businessofgovernment.org/sites/default/files/Using%20Artificial%20Intelligence%20to%20Transform%20Government.pdf.

Pencheva, I., M. Esteve y S.J. Mikhaylov (2018), "Big Data and AI – A transformational shift for government: So, what next for research?", *Public Policy and Administration*, https://doi.org/10.1177/0952076718780537.

Raja, A. (2018), "How will GDPR affect AI?" *Medium*, 30 de octubre, https://medium.com/datadriveninvestor/how-will-gdpr-affect-ai-3f10ed25e4c4.

Rudgard, O. (2019), "Government AI rules to require diverse teams to prevent racist and sexist algorithms", *The Telegraph*, 20 de septiembre, www.telegraph.co.uk/technology/2019/09/20/government-ai-rules-require-diverseteams-prevent-racist-sexist.

Russell, S. y P. Norvig (2016), *Artificial Intelligence: A Modern Approach*, 3rd Edition, Pearson Education, Londres, http://aima.cs.berkeley.edu.

Schneider, T.K. (2019), "Your agency isn't ready for AI", *FCW*, 19 de abril, https://fcw.com/articles/2019/04/19/fcw-perspectives-ai.aspx.

Scholta, H., W. Mertens, M. Kowalkiewicz y J. Becker (2019), "From one-stop shop to no-stop shop: An e-government stage model", *Government Information Quarterly*, Vol. 36/1, pp. 11-26, www.sciencedirect.com/science/article/pii/S0740624X17304239.

Shafique, A. (2018), "Forget jobs. Will robots destroy our public services?" *RSA*, 12 de septiembre, www.thersa.org/discover/publications-and-articles/rsa-blogs/2018/09/forget-jobs.-will-robots-destroy-our-public-services.

Snow, J. (2019), "How artificial intelligence can tackle climate change", *National Geographic*, 18 de julio, www.nationalgeographic.com/environment/2019/07/artificial-intelligence-climate-change.

Soltani, A.A., H. Huang, J. Wu, T.D. Kulkarni y J.B. Tenenbaum (2017), "Synthesizing 3D shapes via modeling multi-view depth maps and silhouettes with deep generative networks", Computer Vision Foundation, http://openaccess.thecvf.com/content_cvpr_2017/papers/Soltani_Synthesizing_3D_Shapes_CVPR_2017_paper.pdf.

Ubaldi, B., E.M., Le Fevre, E. Petrucci, P. Marchionni, C. Biancalana, N. Hiltunen, D.M. Intravaia y C. Yang (2019), "State of the art in the use of emerging technologies in the public sector", *OECD Working Papers on Public Governance*, No. 31, OECD Publishing, París, https://doi.org/10.1787/932780bc-en.

van Ooijen, C., B. Ubaldi y B. Welby (2019), "A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance", *OECD Working Paper*, París, OECD Publishing, https://doi.org/10.1787/09ab162c-en.

Viechnicki, P. y W.D. Eggers (2017), *How much time and money can AI save government? Cognitive technologies could free up hundreds of millions of public sector worker hours*. Deloitte University Press, www2.deloitte.com/content/dam/insights/us/articles/3834_How-much-time-andmoney-can-AI-save-government/DUP_How-much-time-and-money-can-AI-savegovernment.pdf.

Vincent, J. (2019), "Forty percent of 'AI startups' in Europe don't actually use AI, claims report", *The Verge*, 5 de marzo, www.theverge.com/2019/3/5/18251326/ai-startups-europe-fake-40-percent-mmc-report.

Wachter, S., Mittelstadt, B. y L. Floridi (2017), "Why a right to explanation of automated decision-making does not exist in the general data protection regulation", International Data Privacy Law, Vol. 7/2, pp. 76–99, https://doi.org/10.1093/idpl/ipx005.

West, M., R. Kraut y H.E. Chew (2019), *I'd Blush if I Could: Closing Gender Divides in Digital Skills Education*, publicación preparada por la UNESCO para EQUALS Skills Coalition, EQUALS Global Partnership, https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1.

Whittaker, M., K. Crawford, R. Dobbe, G. Fried, E. Kaziunas, V. Mathur, S. Myers West, R. Richardson, J. Schultz y O. Schwartz (2018), *AI Now Report 2018*, AI Now, New York University, Nueva York, https://ainowinstitute.org/AI_Now_2018_Report.pdf.

Wingfield, T., L. Kostopoulos, C. Hodes y N. Miailh (2016), “Artificial Intelligence and the Law of Armed Conflict: Parameters for Discussion”, The Future Society at the Harvard Kennedy School of Government, Cambridge, MA, http://ai-initiative.org/wp-content/uploads/2016/08/AI_MSC.-FINAL.pdf.

Wright, T. (2018), “Canada’s use of artificial intelligence in immigration could lead to break of human rights: study”, *Global News*, 26 de septiembre, https://globalnews.ca/news/4487724/canada-artificial-intelligence-human-rights.

Traducido por





Traducción patrocinada por

